

1971

The mathematical and statistical theory of culling selection

Esmat Moustafa Nouri
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>



Part of the [Statistics and Probability Commons](#)

Recommended Citation

Nouri, Esmat Moustafa, "The mathematical and statistical theory of culling selection " (1971). *Retrospective Theses and Dissertations*. 4417.
<https://lib.dr.iastate.edu/rtd/4417>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

71-21,963

NOURI, Esmat Moustafa, 1933-
THE MATHEMATICAL AND STATISTICAL THEORY OF
CULLING SELECTION.

Iowa State University, Ph.D., 1971
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

**The mathematical and statistical theory
of culling selection**

by

Esmat Moustafa Nouri

**A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY**

Major Subjects: Statistics-Genetics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Heads of Major Departments

Signature was redacted for privacy.

Dean of Graduate College

**Iowa State University
Of Science and Technology
Ames, Iowa**

1971

TABLE OF CONTENTS

	Page
INTRODUCTION	1
NO ENVIRONMENTAL VARIABILITY	9
One Locus - Two Alleles	9
No dominance case: $aa < Aa < AA$	13
Complete dominance case I: $aa < Aa = AA$	24
Complete dominance case II: $aa = Aa < AA$	29
Partial overdominance: $aa < AA < Aa$	32
Complete overdominance: $aa = AA < Aa$	36
Partial underdominance: $Aa < aa < AA$	39
Complete underdominance: $Aa < aa = AA$	42
One Locus - Three Alleles	45
No dominance case with equal effects	46
Complete dominance with equally spaced effects	53
Overdominance	59
Model I: Overdominance with additive effect	59
Model II: Complete overdominance with equal effects	65
Two-Loci Case	68
Two loci with equal effects	72
No dominance at both loci	72
Complete dominance at both loci	100
Overdominance for both loci with interaction	107
Overdominance without interaction	123

	Page
Two loci with unequal effects	131
No dominance at both loci	131
VARIABILITY OF EXPRESSION OF GENETIC FACTORS	144
One-Locus Case	144
Two-Loci Case	170
No dominance at both loci	179
Complete dominance at both loci	185
Overdominance	189
EXTENSION TO n LOCI	201
SUMMARY AND CONCLUSIONS	203
BIBLIOGRAPHY	208
ACKNOWLEDGEMENTS	212
APPENDIX A: FIGURES	213
APPENDIX B: TABLES	242

INTRODUCTION

The theory of genetic selection is an essential part of evolutionary theory if this theory is to progress beyond a purely descriptive account. Different mathematical models regarding selection, natural or artificial, have been considered by several workers. Fisher, Haldane and Wright have made major contributions to the theory of natural selection.

The great bulk of this classical work is directed towards a definite type of selection which may be described as follows. It assumes that there is associated with each genotype a selective value which determines the frequency of zygotes of that genotype that reach adulthood. It then assumes that the adult population which results reproduces by random mating. The most elementary case is that of an infant population segregating at one locus with two alleles A and a with zygotes AA, Aa and aa. Under the assumptions, the zygotic population has Hardy-Weinberg structure. The computation of progress is illustrated by Table 1, where p and q denote the frequency of A and a genes, respectively, and $p+q = 1$.

Table 1. Frequencies and selective values of genotypes.

Genotype	AA	Aa	aa
Frequency	p^2	$2pq$	q^2
Selective value	1	$1-hs$	$1-s$

It is a consequence that the array of adults is

$$\frac{p^2 AA + (1-hs)2pq Aa + (1-s)q^2 aa}{p^2 + (1-hs)2pq + (1-s)q^2}.$$

This then leads to new gene frequencies denoted by p' and q' , where

$$\begin{aligned} p' &= \frac{p^2 + (1-hs)pq}{p^2 + (1-hs)2pq + (1-s)q^2} \\ &= \frac{p(1-hsq)}{1-sq(q+2hp)} \\ q' &= 1 - p' . \end{aligned}$$

Hence the change in gene frequency is

$$\Delta p = p' - p = \frac{[1-h - (1-2h)p]}{1-sq(q+2hp)} spq .$$

This line of development has been pursued vigorously by Haldane (1927), Fisher (1930), Haldane (1931) and Wright (1931). Several other workers, Wright (1942 and 1969), Li (1955 and 1967), and Clark and O'Donald (1964) have presented and utilized the same approach. This approach has been extended to the case of finite population by the use of probability calculus, and descriptions of the present state of knowledge have been given by Moran (1962), Kimura (1965), Wright (1969), and Crow and Kimura (1969), for example. It is worth noting, however, that there are deep difficulties in formulating a mathematical development which incorporates diploidy and mating. The aforementioned book of Moran exhibits the difficulties with great clarity. As a result, most of the finite population theory is partially, or even totally, haploid in nature.

In general, the theory of selection should incorporate the roles of viability, fecundity and mating pattern. An overall theory of selection in genetic population should encompass at least the following aspects:

- (i) differential viability of zygotes of infants to the adult reproductive phase of life;
- (ii) the role of pattern of mating of individuals that are in the reproductive phase;
- (iii) the fecundity of matings, that is, the zygotic output of types of mating;
- (iv) the effects of immigration and emigration;
- (v) the effects of finite size of population;
- (vi) the effects of competition within species;
- (vii) the effects of competition between species.

There has been very extensive work on most of the above aspects, particularly by Wright in the 20's and 30's and by many workers since that period.

It is beyond the scope of the present thesis to make a detailed review and evaluation of this huge body of work. Aspects which seem worthy of comment in the present context are the following. With regard to pattern of mating, the great bulk of the work has utilized a two-phase life, which we may call infancy and adulthood, with random mating of the adults to produce the next generation of infants. The recent work of Karlin and Scudo (1969) is an attempt to incorporate a plausible model of a preferential mating pattern. The past work has used almost exclusively the absence of fecundity differences between matings. Attempts to incorporate such differences were made by Penrose (1949), Bodmer (1965) and by Kempthorne and Pollak (1970), but this was in the context of populations of infinite size. In this work, constant viability and fecundity measures are used. Thus it ignores the requirement that the "selective value" (however it be

defined) of an individual of a particular type depends on its abilities for survival and reproduction relative to that of other individuals.

The role of differential viability of zygotes to adulthood has been studied almost exclusively by the type of model exhibited in Table 1, in which each zygote of particular type has a constant probability of survival to adulthood, which is independent of the nature of the population in which the zygote occurs. This theory is totally applicable, for example, to a consideration of progress of a human population in which certain genotypes are removed from the reproducing population by eugenic measures, the selective value then being the probability that an individual of that genotype is not removed from the reproducing population. In general, however, it seems much more reasonable to suppose that the selective value of a genotype depends on the genotype and on the frequencies of all genotypes in the population.

A theory has been developed over the years for the case of selection for a quantitative attribute. This theory, which dates back to Fisher (1918), is given in most complete form by Griffing (1960), for the consideration of artificial selection--notably culling selection. The basic idea in this development is that there is a scale of merit and all above a truncation point, x , are selected and all below that point are culled. It would seem that this type of selection may be operative in nature and, may, in fact, predominate, the scale of merit being "strength for life" according to some definition of the terms. A more common case, particularly in artificial selection with an economic species, is that there is a scale of merit, and the demands of space and resources lead to the saving of the best S ($0 < S < 1$) of the individuals to reproduce the next generation, and

culling of the remaining fraction $(1-S)$. For example, if an environment can support 1000 adults and 10,000 offspring are produced, then the least fit 90 percent of the offspring have to be culled. If there is natural selection the removal is by death, while if there is artificial selection the removed individuals are those culled by the breeder. This type of selection is the most extensively used method of selection in plant and animal breeding. Whether or not an individual is culled would partially depend on the sorts of individuals he is competing against, so that his "selective value" (however it be defined) would be frequency dependent. For in any generation the worst individuals are culled, and this results in reducing the frequency of the most inferior genotype. It is then possible, after a certain number of generations, for some individuals of another genotype to be culled, although originally all such individuals were allowed to breed. It seems that selection of this general type must be present in natural populations which, like artificial or economic populations, are developing under a definite restriction of the total amount of "living space" available to the population.

This theory of culling selection for a quantitative attribute has utilized intrinsically the model that there is a large number of segregating loci with small effects of the segregating genes. It has, in its present state, the very marked deficiency that it gives only the instantaneous rate of change of the population mean. As selection proceeds, the frequencies of genes change, and the statistical parameters that characterize the genetic variability of the population change. As a result, the long-term changes are not given by the mathematical formulation. There has been a little work on the theory of culling selection for a qualitative

attribute. This type of selection, with no variability of expression of genetic factor and no differences in fecundity, was considered by Van Der Veen (1960), Haldane (1961) and Pollak (1966). They considered the case of one locus with two alleles with assumptions that the population under selection is very large, mating among the selected groups of each generation is at random, generations do not overlap, and that the fraction $(1-S)$ culled in each generation is constant but may be different in the two sexes. It is seen from their findings that the effect of a given intensity of selection, as judged by the fraction culled, falls off after some generations, as is usually the case in practice. Pollak (1966) shows that stable equilibrium can be attained under some cases of selection pressure and that a point of equilibrium does not necessarily correspond to maximum mean genotypic value. For a given situation, the type of equilibrium depended on the relative proportions culled in the two sexes. A major aspect of the present thesis is the study of the progress of infinite populations under the culling or truncation type of selection, in contrast to the study only of equilibria conditions. The selection models considered by Van Der Veen (1960), Haldane (1961) and Pollak (1966) involved a single locus with two alleles, such that the three possible genotypes give invariable phenotypes which are completely distinguishable. It seems desirable to supplement this work with study of the case in which the phenotypes are invariable but not distinct. It is true, in all these cases, that the final or equilibrium status of the population can be derived somewhat easily. However, the actual progress of the population towards its equilibrium conditions is considered to be a highly relevant matter. In other words, the actual response curve over generations to the selection is important.

A deficiency of the above-mentioned work is that the phenotypes associated with the various genotypes are assumed to be absolutely invariable. This condition is one which will be rare in actual biological systems. It is considered desirable, therefore, to develop some theory of the progress of a population in which there is a small number of segregating loci with genetic effects that exhibit variability. It is here that the standard theory of quantitative inheritance breaks down, because as mentioned above the statistical parameters of the population change. The "instantaneous" rate of change of population mean therefore changes also.

It should be noted that the sort of model envisaged includes the possibility of selection for an intermediate on a quantitative attribute, because we may regard the attribute actually under selection as $|\text{phenotype}-T|$ where T is the aimed at intermediate value. In the theory which is discussed in this thesis, the fraction of the population saved is the best S with regard to the attribute under selection. It is necessary to mention this because another type of selection for an intermediate has been considered. This work by Robertson (1956) and Curnow (1964) takes the position that all individuals with quantitative attribute lying in some interval, say, (L,U) are saved. The proportion of the population saved will in this case depend on the genotypic and phenotypic structure of the population.

The aim of the present study is to examine in detail the dynamics of a genetic population under culling selection both without and with environmental variability in the expression on the scale of merit of genotypes and not merely equilibrium conditions or instantaneous rate of change of population mean.

The main difficulty in the development arises because the degree of selection against any one genotype depends critically on the population structure at the time of selection. Hence while the change from one generation to the next can be characterized by "fitness" coefficient for each genotype, for example, w_2 for the AA genotype, w_1 for the Aa genotype, and w_0 for the aa genotype, these "w" values are dependent on the frequency of the genotypes in the population. Thus the work can be regarded as a study of gene frequency dependent selection, in which the actual dependence of the "selective values" of genotypes is determined by the relative positions of the genotypes on a scale of merit and by the relative frequencies in a totally specified way.

NO ENVIRONMENTAL VARIABILITY

One Locus - Two Alleles

Consider a character controlled by a single locus with two alleles, A and a, and denote the three genotypes by AA, Aa and aa. It is assumed in this section that the phenotypes associated with the three genotypes are constant with no variation due to non-genetic causes. It is assumed that reproduction arises by random mating in the population of selected individuals and that there are no fecundity differences between the possible genetic types of mating. It is assumed, also, that the populations are infinite in size so that fluctuations due to sampling will not occur. We shall explore in detail first the case of one locus with two alleles.

Selection applied is by culling with the worst $(1-S)$ individuals culled in every generation and saving of the best $S(0 < S < 1)$ individuals to reproduce the next generation by random mating. With this model, the actual dependence of the "selective values" of genotypes are determined by the relative positions of the genotypes on a scale of merit and by the relative frequencies. The progress of the population under this model depends on the order of merit of the three possible genotypes. Because labelling of the genes is irrelevant, there are seven possible orders that must be considered. The possible orders of the genotypes and the conditions of dominance and selection are given in Table 2.

Table 2. Order of genotypes with conditions of dominance and selection.

Conditions of dominance	Order of genotypes	Conditions of selection
1. No dominance	$aa < Aa < AA$	Selection against a
2. Complete dominance I	$aa < Aa = AA$	Selection against aa
3. Complete dominance II	$aa = Aa < AA$	Selection against a
4. Partial overdominance	$aa < AA < Aa$	Selection against AA, aa
5. Complete overdominance	$aa = AA < Aa$	Selection against AA, aa
6. Partial underdominance	$Aa < aa < AA$	Selection against A and a combination
7. Complete underdominance	$Aa < aa = AA$	Selection against Aa

The progress of the population under selection proceeds in stages. The number of stages of selection depends on the number of distinct possible phenotypes and on the proportion culled in relation to gene frequency. It is clear that with m distinct phenotypes the number of possible stages is m . If the population is in a given stage, the number of possible stages left can be determined. Assume we have m distinct phenotypes. Let the genotypes be arranged in an increasing order depending on their scale of merit and let u_i ($i = 1, 2, \dots, m$) be the frequency of the i -th ordered genotype. Let U_i be the cumulative frequency of the genotypes such that $U_i = \sum_{j=1}^i u_j$. From the definition of U_i we have $U_0 = 0$ and $U_m = 1$. With these notations, the i -th stage is defined by the inequality

$$U_{i-1} \leq (1-S) < U_i \quad i = 1, 2, \dots, m.$$

Then if the population is in the i -th stage, the number of possible stages left is given by $(m-i)$.

In all cases it will be found that there is a mathematically simple relation of p_{n+1} the frequency of one allele, which we take to be the frequency of A, in terms of p_n , the frequency of the same allele in the previous generation. The general procedure that is used, which follows that used extensively by Haldane (1924-27 and later papers) of constructing

$$\begin{aligned}\Delta p_n &= p_{n+1} - p_n \\ &= g(p_n), \text{ say,}\end{aligned}$$

and then replacing this difference equation by the continuous analog

$$\frac{dp_n}{dn} = \Delta p_n = g(p_n) .$$

We shall refer to this procedure by the term continuization, and we shall call the differential equation the continuized analog of the difference equation. This differential equation is then solved to give an approximate relation between n and p_n . The accuracy of this approximating device appears to be extremely difficult to determine by purely mathematical considerations, but there seems to be no recourse but to use this type of approximation. A general obscurity exists with regard to choice of the parameter which one uses in a continuization of the discrete process. For instance, in the above case instead of working with

$$\Delta p_n = g(p_n)$$

and its continuized analog

$$\frac{dp_n}{dn} = g(p_n)$$

we could consider a function $h(p_n) = h_n$ and find

$$\begin{aligned}\Delta h_n &= h(p_{n+1}) - h(p_n) \\ &= H(p_n) \\ &= H[h^{-1}(h_n)] \\ &= k(h_n)\end{aligned}$$

which we could continuize by writing

$$\frac{dh_n}{dn} = k(h_n) .$$

The function of gene frequency $h(p_n)$ may be of the form $\frac{1}{1-p_n}$, $\ln p_n$, or of any other form. Haldane (1927) considered the function $u_n = \frac{p_n}{1-p_n}$.

Accordingly

$$u_{n+1} = \frac{p_{n+1}}{1-p_{n+1}} = U(p_n)$$

and we find

$$\Delta u_n = u_{n+1} - u_n = U(p_n) - u_n ,$$

which is continuized by writing

$$\frac{du_n}{dn} = U(p_n) - u_n .$$

Because of the mathematical difficulty of comparing a priori different models of continuization of the known discrete process, all formulas which

have been derived by a particular continuization have been checked by exact computation on an electronic computer. In general, the continuization which has been used in the present study is the one arises naturally by working with Δp_n expressed as a function of p_n .

No dominance case: $aa < Aa < AA$

Haldane (1961) considered the case of no dominance with selection the same in both sexes. An attempt will be made here to construct a function of gene frequency that changes linearly with generation assuming no differences between the two sexes in the degree of selection. The case when the degree of selection is different in the two sexes will also be considered.

Case 1. The degree of selection is the same in both sexes.

The order of the genotypes is $aa < Aa < AA$. The three genotypes and their frequencies in generation n are as follows:

Genotype	aa	Aa	AA
Frequency	$(1-p_n)^2$	$2p_n(1-p_n)$	p_n^2

We have the following three possible stages of progress.

1. $(1-S) < (1-p_n)^2$

The frequency p of the gene A in the next generation is given by

$$\begin{aligned}
 p_{n+1} &= \frac{1}{S} [p_n^2 + p_n(1-p_n)] \\
 &= \frac{1}{S} p_n .
 \end{aligned}
 \tag{1.1}$$

So, as long as the inequality on S holds,

$$p_n = \frac{1}{S^n} p_0, \quad (1.2)$$

where p_0 is the frequency of A gene in the initial generation. By taking the logarithm of each side of equation (1.2), we find

$$\ln p_n = \ln p_0 - (\ln S)n, \quad (1.3)$$

note that $\ln S$ is negative for $0 < S < 1$. The solution given by (1.3) indicates that $\ln p$ changes linearly with generations with slope given by

$$\beta_1 = - \ln S. \quad (1.4)$$

This function is plotted against generations in Figure 2 for different values of S . In the figure the function used is $\ln p - \ln 2$. The slope of each line is a function of the intensity of selection, the stronger the intensity the higher is the slope.

An attempt will be made here to make an exact comparison of the difference equation Δp_n and its continuous analog $\frac{dp_n}{dn}$. The change in gene frequency in one generation is given by

$$\begin{aligned} \Delta p_n &= p_{n+1} - p_n \\ &= \frac{1}{S} p_n - p_n \\ &= \frac{1-S}{S} p_n. \end{aligned}$$

The change in frequency may be continuized by writing

$$\Delta p_n = \frac{1-S}{S} p_n \doteq \frac{dp_n}{dn} .$$

The above differential equation yields the following solution

$$\ln p_n = \ln p_o + \frac{1-S}{S} n . \quad (1.5)$$

The approximate solution (1.5) indicates that $\ln p_n$ is expected to change linearly with generations with slope given by

$$\hat{\beta} = \frac{1-S}{S} . \quad (1.6)$$

Thus a comparison between the exact solution given by (1.3) and the approximate one given by (1.5) shows that the continuization is appropriate only when selection is weak, i.e. S is close to one. Table 3 gives the values of $(1-S)/S$ and $(-\ln S)$ for different values of S .

Table 3. Comparison between the exact slope $(-\ln S)$ and its estimate $(1-S)/S$ for different values of selection (S).

	S								
	.25	.40	.50	.70	.80	.90	.95	.99	.999
$(1-S)/S$	3.0	1.5	1.0	.43	.25	.111	.05	.0101	.001
$-\ln S$	1.39	.92	.69	.36	.22	.105	.053	.01	.001

From (1.2) it is seen that when S is small the increase in gene frequency due to selection is rapid, and no equilibrium can be reached in this stage. Figure 1 shows the exact change in gene frequency p with generations for different values of S when the initial gene frequency $p_0 = .05$. The increase in gene frequency will continue until $(1-S) \geq (1-p_n)^2$. Replacing p_n by its value from (1.2), we find

$$1-S \geq (1-S^{-n} p_0)^2 . \quad (1.7)$$

Therefore progress according to (1.2) continues until

$$\begin{aligned} n &\geq -\frac{1}{\ln S} [\ln(1 - \sqrt{1-S}) - \ln p_0] \\ &= \frac{1}{\ln S} [\ln p_0 - \ln(1 - \sqrt{1-S})] . \end{aligned}$$

The number of generations required for the first stage can be obtained by taking the square root and then the logarithm of each side

$$n \doteq \frac{1}{\ln S} [\ln p_0 - \ln(1 - \sqrt{1-S})] . \quad (1.8)$$

Table 4 gives the number of generations required for the first stage and the boundary value of p , denoted by p^* , up to which the first stage is operative, for different values of S .

Table 4. Number of generations required for the first stage and boundary gene frequency (p^*).

Initial gene frequency p_0	Selection S					
	.999	.990	.950	.900	.850	.800
.05	3217	285	53	25	15	11
.10	2465	217	40	18	11	8
.15	2024	177	32	14	9	6
.20	1712	148	26	12	7	4
.25	1470	126	22	9	5	3
.50	717	58	9	3	1	1
p^*	.9667	.90	.777	.684	.633	.553

It is seen in Table 4 that when selection is weak, i.e. S is large, the increase in gene frequency will be slow and the first stage will account for the change in frequency for large number of generations, especially when the initial frequency is small. It is seen from (1.2) that the increase in gene frequency will continue to the next stage.

$$2. \quad (1-p_n)^2 \leq (1-S) < 1-p_n^2.$$

In the next generation, we find

$$\begin{aligned}
 p_{n+1} &= \frac{1}{S} \left\{ p_n^2 + \frac{1}{2} [2p_n(1-p_n) - (1-S) + (1-p_n)^2] \right\} \\
 &= \frac{1}{2} + \frac{p_n^2}{2S}.
 \end{aligned} \tag{1.9}$$

The change in gene frequency after one generation of selection is given by

$$\begin{aligned}
\Delta p_n &= p_{n+1} - p_n \\
&= \frac{1}{2S} (S - 2S p_n + p_n^2) \\
&\doteq \frac{dp_n}{dn} .
\end{aligned} \tag{1.10}$$

The solution of equation (1.10) is given by

$$n \doteq 2 \sqrt{\frac{S}{1-S}} \left[\tan^{-1} \sqrt{\frac{S}{1-S}} \left(\frac{p}{S} - 1 \right) \right]_{p_0}^{p_n} . \tag{1.11}$$

By expanding the right-hand side of (1.11) we get

$$\begin{aligned}
n \doteq 2 \sqrt{\frac{S}{1-S}} \sqrt{\frac{S}{1-S}} \left(\frac{p}{S} - 1 \right) - \frac{1}{3} \left[\sqrt{\frac{S}{1-S}} \left(\frac{p}{S} - 1 \right) \right]^3 \\
+ \frac{1}{5} \left[\sqrt{\frac{S}{1-S}} \left(\frac{p}{S} - 1 \right) \right]^5 - \dots \Big|_{p_0}^{p_n} .
\end{aligned} \tag{1.12}$$

For values of p between .5 and 1, the quantity $\frac{p}{S} - 1$ is small when S is large or close to p , and we may write (1.12), approximately,

$$n \doteq 2 \frac{S}{1-S} \left(\frac{p_n}{S} - 1 \right) - 2 \frac{S}{1-S} \left(\frac{p_0}{S} - 1 \right),$$

or

$$p_n \doteq p_0 + \frac{1}{2} (1-S) n. \tag{1.13}$$

Equation (1.13) expresses p_n as a linear function of n with slope given by

$$\hat{\beta}_2 = \frac{1}{2} (1-S) , \tag{1.14}$$

which is a function of S . Figure 2 shows the exact change in p_n with generations for different values of S . It is seen from the figure that linearity does not hold when S is small, and when $S > .5$, the first stage will account for the change for several generations (Table 4).

It is seen from (1.9) that gene frequency will continue to increase until p reaches the value one or $p = \sqrt{S}$.

$$3. \quad (1-S) \geq 1-p_n^2 \text{ or } S < p_n^2.$$

After one generation of selection we find that the population will consist solely of AA individuals and gene frequency at equilibrium is $p = 1$.

Case 2. Degree of selection is different in both sexes.

We will now consider the additive model with selection different in the two sexes. Let S denote the proportion saved from the females and S' from the males with $S > S'$. The females in the n -th generation produce the gametic array $p_n A + (1-p_n)a$, while the males of that generation produce the gametic array $q_n A + (1-q_n)a$. Then if the survivors mate at random the genotypic array at birth in the next generation is

$$p_n q_n AA + (p_n + q_n - 2p_n q_n)Aa + (1-p_n)(1-q_n)aa.$$

There are five possible stages of progress.

$$1. \quad (1-S) < (1-S') < (1-p_n)(1-q_n).$$

The frequencies in the next generation are given by

$$\begin{aligned}
p_{n+1} &= \frac{1}{S} \{p_n q_n + \frac{1}{2}(p_n + q_n - 2p_n q_n)\} \\
&= \frac{1}{2S} (p_n + q_n) \\
q_{n+1} &= \frac{1}{S'} \{p_n q_n + \frac{1}{2} (p_n + q_n - 2p_n q_n)\} \\
&= \frac{1}{2S'} (p_n + q_n). \tag{1.15}
\end{aligned}$$

Equations (1.15) indicate that, under this condition, the gene ratio

$$\frac{p_n}{q_n} = \frac{S'}{S}$$

is constant over generations during the first stage. So, as long as the inequality on S and S' holds,

$$\begin{aligned}
p_{n+1} &= \left(\frac{S+S'}{2SS'}\right)^n \left(\frac{p_o + q_o}{2S}\right) \\
q_{n+1} &= \left(\frac{S+S'}{2SS'}\right)^n \left(\frac{p_o + q_o}{2S'}\right). \tag{1.16}
\end{aligned}$$

For small values of $(1-S)$ and $(1-S')$, equation (1.16) takes the approximate form

$$\begin{aligned}
p_{n+1} &= \left[\frac{1}{2} (S+S')\right]^n \frac{1}{2} (2-S) (p_o + q_o) \\
q_{n+1} &= \left[\frac{1}{2} (S+S')\right]^n \frac{1}{2} (2-S') (p_o + q_o).
\end{aligned}$$

Equations (1.16) show that the gene frequencies will continue to increase with generations and the increase is rapid when S' is small. The increase will continue until

$$(1-S') \geq (1-p_n)(1-q_n).$$

Using the values for p_n and q_n given by (1.16) we get

$$(1-S') \geq \left\{1 - \left(\frac{S+S'}{2SS'}\right)^{n-1} \left(\frac{p_o + q_o}{2S}\right)\right\} \left\{1 - \left(\frac{S+S'}{2SS'}\right)^{n-1} \left(\frac{p_o + q_o}{2S'}\right)\right\}.$$

It is seen that the above inequality is equivalent to

$$\left(\frac{S+S'}{2SS'}\right)^{n-1} (p_o + q_o) \geq (S+S') - \sqrt{4c' - 3(c')^2 - 6cc' + c^2 + 4c(c')^2}, \quad (1.17)$$

where $c = 1-S$ and $c' = 1-S'$. If all the females are saved, i.e. $S = 1$, the inequality given by 1.17 reduces to

$$\left(\frac{1+S'}{2S'}\right)^{n-1} (p_o + q_o) \geq (1+S') - \sqrt{(1-S')(1+3S')}. \quad (1.18)$$

On taking logarithms of both sides, the smallest number of generations which satisfies (1.17) or (1.18) can be found. From (1.17) we have

$$n \doteq 1 + \frac{\ln[(S+S') - \sqrt{4c' - 3(c')^2 - 6cc' + c^2 + 4c(c')^2}] - \ln(p_o + q_o)}{\ln(S+S') - \ln 2SS'} \quad (1.19)$$

and using (1.18) we get

$$n \doteq 1 + \frac{\ln[(1+S') - \sqrt{(1-S')(1+3S')}] - \ln(p_o + q_o)}{\ln(1+S') - \ln 2S'}, \quad (1.20)$$

where $c = 1-S$ and $c' = 1-S'$.

Using (1.19), the number of generations required for the first stage when $p_0 = q_0 = .05$ is given in Table 5 for different values of S and S' .

Table 5. Number of generations required for different values of S and S' .

		S'					
		.999	.95	.9	.75	.5	.25
S	1.0	5922	106	50	15	5	2
	.95		54	49	13	3	2
	.90			25	12	2	1
	.75				8	1	1
	.50					1	1

For small values of S and S' the number required is small and it increases with S and S' . The change in gene frequencies will continue to the next stage.

$$2. \quad (1-S) < (1-p_n)(1-q_n) < (1-S') < 1 - p_n q_n.$$

We find

$$\begin{aligned} p_{n+1} &= \frac{p_n + q_n}{2S} \\ q_{n+1} &= \frac{S' + p_n q_n}{2S'} . \end{aligned} \quad (1.21)$$

The change in the gene frequencies is given by

$$\begin{aligned} \Delta p_n &= p_{n+1} - p_n = \frac{1}{2S} \{ p_n (1 - 2S) + q_n \} \\ \Delta q_n &= q_{n+1} - q_n = \frac{1}{2} - \left(1 - \frac{p_n}{2S'} \right) q_n . \end{aligned} \quad (1.22)$$

Under the present additive model the change in the gene frequencies will continue to the next stage.

$$3. \quad (1-p_n)(1-q_n) < (1-S) < (1-S') < 1 - p_n q_n.$$

In this stage we find

$$\begin{aligned} p_{n+1} &= \frac{1}{2} + \frac{p_n q_n}{2S} \\ q_{n+1} &= \frac{1}{2} + \frac{p_n q_n}{2S'} . \end{aligned} \quad (1.23)$$

The change in the gene frequencies due to one generation of selection is given by

$$\begin{aligned} \Delta p_n &= p_{n+1} - p_n = \frac{1}{2} + p_n \left(\frac{q_n}{2S} - 1 \right) \\ \Delta q_n &= q_{n+1} - q_n = \frac{1}{2} + q_n \left(\frac{p_n}{2S'} - 1 \right) . \end{aligned} \quad (1.24)$$

It is seen from (1.23) that the ratio

$$\frac{p_n - \frac{1}{2}}{q_n - \frac{1}{2}} = \frac{S'}{S} ,$$

is constant over generations during this stage. The change in the gene frequencies continues to the next stage.

$$4. \quad (1-p_n)(1-q_n) < (1-S) < 1 - p_n q_n < (1-S').$$

Since only AA individuals are allowed to reproduce among the males, we find

$$q_{n+1} = 1$$

$$p_{n+1} = \frac{1}{2} + \frac{p_n q_n}{2S}$$

$$p_{n+2} = \frac{1}{2S'} [S' + p_{n+1}] . \quad (1.25)$$

At equilibrium we have $p = 1$ and $q = 1$.

$$5. \quad (1-S') > (1-S) > 1 - p_n q_n.$$

It is clear in this stage that after one generation of selection we find $p = 1$ and $q = 1$, which are the equilibrium values.

Complete dominance case I: $aa < Aa = AA$

The order of the three genotypes is $aa < Aa = AA$. Haldane (1961) considered this model with selection different in the two sexes and $S > S'$, where S and S' are the proportions saved from the females and the males, respectively. Since the two genotypes Aa and AA are not distinguishable from each other, we find only two possible stages of progress.

$$1. \quad (1-S) < (1-p_n)^2.$$

We find

$$\begin{aligned} p_{n+1} &= \frac{1}{S} [p_n^2 + p_n(1-p_n)] \\ &= \frac{1}{S} p_n . \end{aligned} \quad (1.26)$$

So, as long as the inequality on S holds, we find

$$p_n = S^{-n} p_0 . \quad (1.27)$$

Equation (1.27) is the same as (1.2) obtained under the first stage for the model with no dominance. Thus equations (1.1) to (1.8) obtained above apply to this stage.

We consider here the change in the mean of the population. Let μ_{00} , μ_{10} and μ_{11} be the mean phenotypic values of the genotypes aa, Aa and AA, respectively. The mean of the population in generation n is given by

$$W_n = (1-p_n)^2 \mu_{00} + 2p_n(1-p_n) \mu_{10} + p_n^2 \mu_{11} . \quad (1.28)$$

Under the complete dominance model we may take $\mu_{00} = 0$ and $\mu_{10} = \mu_{11} = 1$. Consequently, (1.28) reduces to

$$W_n = 2p_n(1-p_n) + p_n^2 = 2p_n - p_n^2 . \quad (1.29)$$

Replacing p_n by its value from (1.27) we find

$$\begin{aligned} W_n &= 2S^{-n} p_o - S^{-2n} p_o^2 \\ &= S^{-n} p_o (2 - S^{-n} p_o) \\ &= W_o + [2(S^{-n} - 1) - (S^{-2n} - 1)p_o] p_o , \end{aligned} \quad (1.30)$$

where $W_o = 2p_o - p_o^2$. It is seen from (1.30) that the mean of the population will continue to increase with generations. Table 4 gives the number of generations required for the first stage and the boundary value of p, denoted by p^* , up to which the first stage is operative, for different values of S. The first stage will continue until

$$(1-S) \geq (1-p_n)^2 = 1 - (2p_n - p_n^2) = 1-W_n.$$

Thus the boundary value of W , denoted by W^* is given by S .

$$2. \quad (1-S) \geq (1-p_n)^2.$$

We find

$$\begin{aligned} p_{n+1} &= \frac{p_n^2 + p_n(1-p_n)}{p_n^2 + 2p_n(1-p_n)} \\ &= \frac{1}{2-p_n}. \end{aligned} \quad (1.31)$$

Equation (1.31) shows that the effect of selection during the second stage becomes very slow and is never complete. Figure 1 shows the exact change in gene frequency p with generations for different values of S . It can be shown that the solution of (1.31) is

$$p_n = \frac{n - (n-1)p_o}{(n+1) - np_o} = \frac{n(1-p_o) + p_o}{n(1-p_o) + 1}. \quad (1.32)$$

The above solution (1.32) may be written in the form

$$n = \frac{p_n - p_o}{(1-p_o)(1-p_n)} = \frac{1}{1-p_n} - \frac{1}{1-p_o} \quad (1.33)$$

The solution given by (1.32) can be proved by induction. First the solution is true for $n = 1$, for

$$p_1 = \frac{1}{2-p_o}$$

which is the form given by (1.31). Assume the relation given by (1.32) is true for $n = k$

$$p_k = \frac{k - (k-1)p_0}{(k+1) - kp_0} . \quad (1.34)$$

For $n = k+1$ we find, using (1.31),

$$p_{k+1} = \frac{1}{2-p_k} ,$$

which by (1.34) becomes

$$p_{k+1} = \frac{1}{2 - \frac{k - (k-1)p_0}{(k+1) - kp_0}} = \frac{(k+1) - [(k+1) - 1]p_0}{[(k+1) + 1] - (k+1)p_0} .$$

Hence the relation is true for $n = k+1$.

It is seen from the exact solution given by (1.32) that p_n approaches unity as n gets large.

The change in gene frequency in one generation is given by

$$\begin{aligned} \Delta p_n &= p_{n+1} - p_n \\ &= \frac{1}{2-p_n} - p_n \\ &= \frac{(1-p_n)^2}{2-p_n} . \end{aligned} \quad (1.35)$$

Equilibrium exists when $\Delta p_n = 0$. It is seen from (1.35) that at equilibrium we have $p = 1$.

An attempt will be made here to make an exact comparison between the continuous time model, $\frac{dp_n}{dn}$, and the discrete time model, Δp_n . The change in gene frequency given by (1.35) may be continuized by writing

$$\Delta p_n = \frac{(1-p_n)}{2-p_n} \doteq \frac{dp_n}{dn}.$$

The above differential equation yields the following relation

$$n = \frac{1}{1-p_n} - \frac{1}{1-p_0} - \ln \frac{1-p_n}{1-p_0}. \quad (1.36)$$

Thus a comparison between the exact solution given by (1.33) and the approximate solution given by (1.36) shows that the two differ by the term $\ln \frac{1-p_n}{1-p_0}$. Thus the two solutions will yield the same result only if $\ln \frac{1-p_n}{1-p_0} = 0$ or $p_n = p_0$. This suggests that gene frequency is not changing or the factors influencing gene frequency are not operating.

Equation (1.36) indicates that the function of the gene frequency $\frac{1}{1-p_n} - \ln(1-p_n)$ is expected to change linearly with generations with slope given by

$$\hat{\beta}_2 = 1, \quad (1.37)$$

which is independent of selection. Figure 2 shows the exact change in this function with generations for different values of S . The function used in Figure 2 is $\frac{1}{1-p_n} - \ln(1-p_n) - 2.69$. It is seen from the figure that the lines for different degrees of selection are approximately parallel, indicating that the change in gene frequency during this stage is independent of selection.

Table 2, Appendix B, gives the values for the exact slope β and their estimates $\hat{\beta}$ for different values of S . The β 's are estimated from the linear relationship presented in Figure 2. The $\hat{\beta}$'s are the estimates of the slope obtained from the derived equations for each stage. It is seen from the table that an approximate linearity holds for each stage and that the $\hat{\beta}$'s give close estimate of the β 's.

We now consider the change in the mean of the population during the second stage. Using the exact solution for p_n given by (1.32), the mean of the population given by (1.29) is given by

$$\begin{aligned} W_n &= 2 \left[\frac{n(1-p_o) + p_o}{n(1-p_o) + 1} \right] - \left[\frac{n(1-p_o) + p_o}{n(1-p_o) + 1} \right]^2 \\ &= W_o + \frac{n(1-p_o)^3}{[n(1-p_o) + 1]^2} [2 + n(1-p_o)] , \end{aligned} \quad (1.38)$$

where $W_o = 2p_o - p_o^2$ is the mean of the initial population. It is seen from (1.38) the mean of the population will approach unity as n increases. It can also be shown that the maximum value for W under the present model is one and is attained when p reaches the equilibrium value of one.

Complete dominance case II: $aa = Aa < AA$

The order of the genotypes is $aa = Aa < AA$. Under this model, the A gene is superior to the a gene but the recessive gene may act as an inhibitor to A when it is present, or the a gene could be lethal. In this case there are two possible stages of progress.

$$1. \quad (1-S) < (1-p_n)^2 + 2p_n(1-p_n) = 1-p_n^2.$$

In this case not all the aa's and Aa's are culled in one generation.

The array among the culled individuals in generation n is

$$\frac{(1-p_n)^2 aa + 2p_n(1-p_n)Aa}{(1-p_n)^2 + 2p_n(1-p_n)} = \frac{(1-p_n)aa + 2p_n Aa}{1 + p_n}$$

The array among the saved individuals is

$$\begin{aligned} \frac{1}{S} \{ p_n^2 AA + [2p_n(1-p_n) - (1-S) \frac{2p_n}{1+p_n}] Aa \\ + [(1-p_n)^2 - (1-S) \frac{1-p_n}{1+p_n}] aa \} \end{aligned}$$

Thus the new gene frequency is

$$\begin{aligned} p_{n+1} &= \frac{1}{S} \{ p_n^2 + \frac{1}{2} [2p_n(1-p_n) - (1-S) \frac{2p_n}{1+p_n}] \} \\ &= \frac{1}{S} \frac{p_n(S+p_n)}{1+p_n} \end{aligned} \quad (1.39)$$

The change in gene frequency in one generation is given by

$$\Delta p_n = p_{n+1} - p_n.$$

Using (1.39), we find

$$\Delta p_n = \frac{1-S}{S} \frac{p_n^2}{1+p_n}. \quad (1.40)$$

The change in frequency may be continuized by writing

$$\Delta p_n = \frac{1-S}{S} \frac{p_n^2}{1+p_n} \doteq \frac{dp_n}{dn} . \quad (1.41)$$

The above differential equation yields the following solution

$$\frac{1}{p_n} - \ln p_n = \left(\frac{1}{p_0} - \ln p_0 \right) + \frac{1-S}{S} n . \quad (1.42)$$

Thus the left-hand side of (1.42) is expected to change linearly with generations with slope given by $(1-S)/S$.

We consider now the change in the mean of the population during this stage. The mean of the population in generation n is given by

$$W_n = (1-p_n)^2 \mu_{00} + 2p_n(1-p_n) \mu_{10} + p_n^2 \mu_{11} ,$$

where μ_{00} , μ_{10} and μ_{11} are the mean phenotypic values of the genotypes aa , Aa and AA , respectively. For the present model we may take $\mu_{00} = \mu_{10} = 0$ and $\mu_{11} = 1$. Accordingly, we find

$$W_n = p_n^2 . \quad (1.43)$$

The change in the population mean is given by

$$\begin{aligned} \Delta W_n &= W_{n+1} - W_n \\ &= p_{n+1}^2 - p_n^2 \\ &= \Delta p_n (\Delta p_n + 2p_n) , \end{aligned} \quad (1.44)$$

where $\Delta p_n = p_{n+1} - p_n$. Replacing Δp_n by its value given by (1.40) we find

$$\Delta W_n = \frac{1-S}{S} \frac{p_n^2}{1+p_n} \frac{[2S + (1+S)p_n]p_n}{S(1+p_n)} . \quad (1.45)$$

From (1.39) and (1.40) it is seen that gene frequency as well as the mean of the population will continue to increase up to the next stage.

$$2. \quad (1-S) \geq 1 - p_n^2.$$

After one generation of selection the population will consist solely of AA individuals and we have at equilibrium $p = 1$. It is seen from (1.43) that the equilibrium point $p = 1$ corresponds to a maximum for the mean of the population.

Partial overdominance: $aa < AA < Aa$

Pollak (1966) has examined in detail the case in which the order of merit of the three genotypes is $aa < AA < Aa$. He assumed different degrees of culling in the males and females and that selection is stronger in the males than in the females, which is reasonable for most artificial animal breeding plans. Under this case there are five possible stages in the selection process. Denoting by $(1-S)$ the proportion culled from the females and $(1-S')$ the proportion culled from the males, Pollak showed that the progress of the population proceeds in sequences depending on the initial conditions. For some stages where equilibrium is possible, the conditions for the equilibrium and the approach to it depend on the relative magnitude of selection intensities.

We assume here a more simple case in which there are no differences between the sexes with regard to the degree of selection, i.e. $S = S'$. The genotypes and their frequencies in generation n are as follows:

Genotype	aa	AA	Aa
Frequency	$(1-p_n)^2$	p_n^2	$2p_n(1-p_n)$

Under these conditions there are three possible stages of progress.

$$1. \quad (1-S) < (1-p_n)^2.$$

The frequency p of the A gene in the next generation is given by

$$\begin{aligned} p_{n+1} &= \frac{1}{S} [p_n^2 + p_n(1-p_n)] \\ &= \frac{1}{S} p_n \end{aligned} \quad (1.46)$$

So, as long as the inequality on S holds,

$$p_n = \frac{1}{S^n} p_0, \quad (1.47)$$

where p_0 is the frequency of A gene in the initial population. This stage is the same as the first stage discussed above under the model of no dominance, and equations (1.1) to (1.8) apply to this stage.

$$2. \quad (1-p_n)^2 \leq (1-S) < (1-p_n)^2 + p_n^2 = 1 - 2p_n(1-p_n).$$

In this stage we have in the next generation

$$\begin{aligned} p_{n+1} &= \frac{1}{S} \{p_n(1-p_n) + [p_n^2 - (1-S) + (1-p_n)^2]\} \\ &= \frac{1}{S} [S - p_n(1-p_n)]. \end{aligned} \quad (1.48)$$

The change in gene frequency in one generation is given by

$$\begin{aligned}\Delta p_n &= p_{n+1} - p_n \\ &= \frac{1}{S} (1-p_n)(S-p_n) .\end{aligned}\quad (1.49)$$

At equilibrium we have $\Delta p_n = 0$, which is satisfied in this stage by the equilibrium value $p = S$. In order for an equilibrium to take place in this stage, the equilibrium value $p = S$ must satisfy the inequality on S which defines the present stage. Hence we have the equilibrium value

$$p = S \quad \text{if } S > \frac{1}{2} . \quad (1.50)$$

If S is small, i.e. $S \leq \frac{1}{2}$, the two homozygotes will be culled and we have at equilibrium

$$p = \frac{1}{2} \quad \text{if } S \leq \frac{1}{2} . \quad (1.51)$$

Using $S = 1-c$, Pollak (1966) shows that the equilibrium values given by (1.50) and (1.51) are stable.

The change in gene frequency may be continuized by writing

$$\Delta p_n \doteq \frac{dp_n}{dn} .$$

Using (1.34), we find

$$\frac{dp_n}{dn} \doteq \frac{1}{S} (S-p_n)(1-p_n) .$$

The above differential equation has the solution

$$\ln \frac{1-p_n}{S-p_n} = \ln \frac{1-p_o}{S-p_o} + \frac{1-S}{S} n, \quad (1.52)$$

or

$$p_n = \frac{1 - S \left(\frac{1-p_o}{S-p_o} \right) \exp\left(\frac{1-S}{S}n\right)}{1 - \left(\frac{1-p_o}{S-p_o} \right) \exp\left(\frac{1-S}{S}n\right)}. \quad (1.53)$$

Equation (1.52) defines a function of gene frequency which is expected to change linearly with generations with slope given by

$$\hat{\beta} = \frac{1-S}{S}, \quad (1.54)$$

depending on S . Figure 2 shows the exact change in the function given by (1.52) with generations for $S = .75$. When $S > .75$, it is seen from Table 4 that the first stage will account for the change in frequency for a large number of generations. The function used in Figure 2 is $\ln \frac{1-p_n}{S-p_n} - \ln \frac{.5}{S-.5}$.

$$3. \quad (1-S) \geq (1-p_n)^2 + p_n^2.$$

It is clear in this stage that after one generation of selection the adult population will consist solely of Aa individuals. Thus we have the equilibrium point $p = \frac{1}{2}$ and is stable. From the defining inequality for this stage, it is seen that for the equilibrium value $p = \frac{1}{2}$ to take place in this stage we must have $S \leq \frac{1}{2}$.

Complete overdominance: $aa = AA < Aa$

The order of the genotypes is $aa = AA < Aa$. There are two possible stages of progress.

$$1. \quad (1-S) < (1-p_n)^2 + p_n^2.$$

In this case not all the aa and AA individuals are culled in one generation. The array among the culled individuals in generation n is

$$\frac{(1-p_n)^2 aa + p_n^2 AA}{(1-p_n)^2 + p_n^2} = \frac{(1-p_n)^2 aa + p_n^2 AA}{1 - 2p_n + 2p_n^2}.$$

The array among the saved individuals in the same generation is

$$\begin{aligned} & \frac{1}{S} \left\{ [p_n^2 - (1-S) \frac{p_n^2}{1 - 2p_n + 2p_n^2}] AA + 2p_n(1-p_n)Aa \right. \\ & \left. + [(1-p_n)^2 - (1-S) \frac{(1-p_n)^2}{1 - 2p_n + 2p_n^2}] aa \right\}. \end{aligned}$$

Thus the new gene frequency is

$$\begin{aligned} p_{n+1} &= \frac{1}{S} \left\{ [p_n^2 - (1-S) \frac{p_n^2}{1 - 2p_n + 2p_n^2}] + p_n(1-p_n) \right\} \\ &= \frac{1}{S} [1 - (3-S)p_n + 2p_n^2] \frac{p_n}{1 - 2p_n + 2p_n^2}. \end{aligned} \quad (1.55)$$

The change in gene frequency in one generation is given by

$$\Delta p_n = p_{n+1} - p_n .$$

Using (1.55), we find

$$\Delta p_n = \frac{(1-S)}{S} \frac{p_n (1-p_n) (1-2p_n)}{1 - 2p_n + 2p_n^2} . \quad (1.56)$$

Equilibrium exists when $\Delta p_n = 0$. Using (1.56), the equilibrium values are 0, $\frac{1}{2}$ and 1. For the defining inequality for this stage and in order for the non-trivial equilibrium given by $p = \frac{1}{2}$ to take place in this stage we must have $S > \frac{1}{2}$.

The change in gene frequency may be continuized by writing (1.56) in the form

$$\Delta p_n = \frac{1-S}{S} \frac{p_n (1-p_n) (1-2p_n)}{1 - 2p_n + 2p_n^2} \doteq \frac{dp_n}{dn} .$$

The above equation has the following solution

$$\ln p_n - 2 \ln \left| \frac{2p_n - 2}{2p_n - 1} \right| = \ln p_o - 2 \ln \left| \frac{2p_o - 2}{2p_o - 1} \right| + \frac{1-S}{S} n . \quad (1.57)$$

Thus the function of gene frequency given by the left-hand side of (1.57) is expected to change linearly with generations with slope given by $(1-S)/S$.

We consider now the change in the mean of the population during this stage. The population mean in generation n is given by

$$W_n = (1-p_n)^2 \mu_{00} + 2p_n(1-p_n)\mu_{10} + p_n^2 \mu_{11},$$

where μ_{00} , μ_{10} and μ_{11} are the mean phenotypic values of the genotypes aa, Aa, AA, respectively. For the present model we may take $\mu_{00} = \mu_{11} = 0$ and $\mu_{10} = 1$. Accordingly, we find

$$W_n = 2p_n(1-p_n). \quad (1.58)$$

The change in the mean in one generation is given by

$$\begin{aligned} \Delta W_n &= W_{n+1} - W_n \\ &= 2p_{n+1}(1-p_{n+1}) - 2p_n(1-p_n) \\ &= 2\Delta p_n(1-\Delta p_n - 2p_n), \end{aligned} \quad (1.59)$$

where $\Delta p_n = p_{n+1} - p_n$. Replacing Δp_n by its value given by (1.56) we find

$$\Delta W_n = \frac{2(1-S)}{S^2} \frac{p_n(1-p_n)(1-2p_n)^2(S-2p_n+2p_n^2)}{1-2p_n+2p_n^2}. \quad (1.60)$$

It is seen from (1.58) that the maximum value of W is attained when $p = \frac{1}{2}$ at which value $W = \frac{1}{2}$ and $\Delta W = 0$.

$$2. \quad (1-S) \geq (1-p_n)^2 + p_n^2.$$

After one generation the population will consist solely of Aa individuals and we have at equilibrium $p = \frac{1}{2}$. For the defining inequality for this stage, in order for the equilibrium $p = \frac{1}{2}$ to take place in this

stage we must have $S \leq \frac{1}{2}$. This stage is the same as the third stage discussed above for the partial overdominance case. It is seen from (1.58) that the mean of the population will reach its maximum value at the equilibrium point $p = \frac{1}{2}$, in which case we have $W = \frac{1}{2}$ and $\Delta W = 0$.

Partial underdominance: $Aa < aa < AA$

The order of the genotypes is $Aa < aa < AA$. Accordingly, we have three possible stages of progress.

1. $(1-S) < 2p_n(1-p_n)$.

We find

$$\begin{aligned} p_{n+1} &= \frac{1}{S} \left\{ p_n^2 + \frac{1}{2} [2p_n(1-p_n) - (1-S)] \right\} \\ &= \frac{1}{S} \left[p_n - \frac{1}{2} (1-S) \right]. \end{aligned} \quad (1.61)$$

So, as long as the inequality on S holds,

$$p_n = \frac{1}{S^n} p_0 - \frac{1-S^n}{2S^n}. \quad (1.62)$$

Equation (1.62) defines gene frequency p_n in terms of n , S and p_0 . Using (1.62), we may write

$$p_n = \frac{1}{S^n} \left(p_0 - \frac{1}{2} \right) + \frac{1}{2}, \quad (1.63)$$

or

$$p_n - \frac{1}{2} = \frac{1}{S^n} \left(p_0 - \frac{1}{2} \right).$$

By taking the logarithm of each side we find

$$\begin{aligned} \ln(p_n - \frac{1}{2}) &= \ln(p_o - \frac{1}{2}) - (\ln S)n \quad \text{if } p_o > \frac{1}{2} \\ \ln(\frac{1}{2} - p_n) &= \ln(\frac{1}{2} - p_o) + (\ln S)n \quad \text{if } p_o < \frac{1}{2}. \end{aligned} \quad (1.64)$$

Equation (1.64) defines a function of the gene frequency which changes linearly with generations with slope given by

$$\hat{\beta} = - \ln S, \quad (1.65)$$

which is a function of S .

The change in gene frequency in one generation is given by

$$\Delta p_n = p_{n+1} - p_n.$$

Using (1.61), we find

$$\Delta p_n = \frac{1-S}{S} (p_n - \frac{1}{2}). \quad (1.66)$$

Equilibrium exists when $p_n = 0$. Using (1.66), we find the equilibrium value $p = \frac{1}{2}$. For equilibrium point $p = \frac{1}{2}$ to take place in this stage we must have $(1-S) < 2p(1-p)$ or $S > \frac{1}{2}$.

The change in gene frequency will continue until $(1-S) \geq 2p_n(1-p_n)$.

Using (1.62), we find

$$(1-S) \geq 2 \left(\frac{p_o}{S^n} - \frac{1-S^n}{2S^n} \right) \left(1 - \frac{p_o}{S^n} + \frac{1-S^n}{2S^n} \right) \quad (1.67)$$

The above inequality may be written in the form

$$(1-S) \geq \frac{1}{2} \left[1 - \frac{(1-2p_o)^2}{S^{2n}} \right] . \quad (1.68)$$

Therefore progress according to (1.62) continues until

$$n \leq \frac{1}{2\ln S} [2\ln(1-2p_o) - \ln(2S-1)] , \quad (1.69)$$

which is valid only if $S > \frac{1}{2}$.

$$2. \quad 2p_n(1-p_n) \leq (1-S) < 1 - p_n^2.$$

We find

$$p_{n+1} = \frac{1}{S} p_n^2 . \quad (1.70)$$

So, as long as the inequality on S holds,

$$p_n = \frac{1}{S^{2^n-1}} p_o^{2^n} . \quad (1.71)$$

Equation (1.71) may be written in the form

$$\ln[\ln(\frac{p_n}{S})] = \ln[\ln(\frac{p_o}{S})] + (\ln 2)n . \quad (1.72)$$

Equation (1.72) defines a function of gene frequency which changes linearly with generations with slope given by $\ln 2$, independent of selection.

The change in gene frequency is given by

$$\Delta p_n = p_{n+1} - p_n .$$

Using equation (1.49), we find

$$\Delta p_n = \frac{1}{S} p_n (p_n - S) . \quad (1.73)$$

Equilibrium exists when $\Delta p_n = 0$. Using (1.73), it is seen that equilibrium exists in this stage for

$$p = S . \quad (1.74)$$

For the equilibrium point $p = S$ to take place in this stage we must have $2p(1-p) < 1-S < 1-p^2$ or $S < \frac{1}{2}$.

$$3. \quad (1-S) \geq 1 - p_n^2 .$$

After one generation of selection the population will consist solely of AA individuals and we have at equilibrium $p = 1$.

Complete underdominance: $Aa < aa = AA$

The order of the genotypes is $Aa < aa = AA$. There are two possible stages.

$$1. \quad (1-S) < 2p_n(1-p_n) .$$

This stage is the same as the first stage discussed above under the partial underdominance model. Accordingly, we have

$$p_{n+1} = \frac{1}{S} \left[p_n - \frac{1}{2} (1-S) \right] \quad (1.75)$$

$$\begin{aligned} p_n &= \frac{1}{S^n} p_o - \frac{1-S^n}{2S^n} \\ &= \frac{1}{2} + \frac{1}{S^n} \left(p_o - \frac{1}{2} \right) . \end{aligned} \quad (1.76)$$

Equations (1.64) to (1.69) also apply to this stage.

We now consider the change in the mean of the population during this stage. The population mean in generation n is given by

$$W_n = (1-p_n)^2 \mu_{00} + 2p_n(1-p_n) \mu_{10} + p_n^2 \mu_{11} ,$$

where μ_{00} , μ_{10} and μ_{11} are the mean phenotypic values of the genotypes aa , Aa and AA , respectively. For the present model we may take $\mu_{10} = 0$ and $\mu_{00} = \mu_{11} = 1$. Accordingly, we have

$$W_n = (1-p_n)^2 + p_n^2 = 1 - 2p_n + 2p_n^2 . \quad (1.77)$$

Replacing p_n by its value given by (1.76) we find

$$W_n = \frac{1}{2} + 2 s^{-2n} \left(p_0 - \frac{1}{2}\right)^2 . \quad (1.78)$$

It is seen from (1.77) that the maximum value of W does not occur for $p = \frac{1}{2}$ but for values of p near zero or one.

$$2. \quad (1-s) \geq 2p_n(1-p_n).$$

We find

$$p_{n+1} = \frac{p_n^2}{p_n^2 + (1-p_n)^2} . \quad (1.79)$$

The change in gene frequency is given by

$$\begin{aligned}\Delta p_n &= p_{n+1} - p_n \\ &= - \frac{p_n(1-p_n)(1-2p_n)}{1 - 2p_n + 2p_n^2} .\end{aligned}\quad (1.80)$$

Equilibrium exists when $p_n = 0$. From (1.80) the equilibrium values are 0 , $\frac{1}{2}$ and 1 . From (1.80) it is seen that in order for the equilibrium value of $\frac{1}{2}$ to occur in this stage we must have $S \leq \frac{1}{2}$.

The change in gene frequency during this stage may be continuized by writing (1.80) in the form

$$\Delta p_n = - \frac{p_n(1-p_n)(1-2p_n)}{1 - 2p_n + 2p_n^2} \doteq \frac{dp_n}{dn} .$$

The above differential equation yields the following solution

$$\ln p_n + \ln \left| \frac{2p_n - 2}{2p_n - 1} \right| = \ln p_o + \ln \left| \frac{2p_o - 2}{2p_o - 1} \right| - n. \quad (1.81)$$

Thus the function of the gene frequency given by the left-hand side of (1.81) is expected to change linearly with generations with slope given by (-1) .

The change in the population mean W during this stage is given by

$$\Delta W_n = W_{n+1} - W_n .$$

Using (1.77), we find

$$\Delta W_n = 2\Delta p_n (2p_n - 2 + \Delta p_n) ,$$

where Δp_n is given by (1.80). Replacing Δp_n by its value given by (1.80) we find

$$\Delta W_n = \frac{2p_n(1-p_n)(1-2p_n)(2-5p_n+3p_n^2)}{(1-2p_n+2p_n^2)^2} . \quad (1.82)$$

Thus the non-trivial equilibrium $p = \frac{1}{2}$ for $S \leq \frac{1}{2}$ will result in $\Delta p = \Delta W = 0$ and it does not correspond to a maximum for W .

One Locus - Three Alleles

We consider now the case of one locus with three alleles A_1 , A_2 and A_3 with frequencies denoted by p_1 , p_2 and p_3 , respectively, and $p_1 + p_2 + p_3 = 1$. The mean phenotypic values and the frequencies of the six possible genotypes in generation n are presented in Table 6.

Table 6. Mean phenotypic values and frequencies of genotypes.

Genotype	A_1A_1	A_1A_2	A_1A_3	A_2A_2	A_2A_3	A_3A_3
Frequency	$p_{1,n}^2$	$2p_{1,n}p_{2,n}$	$2p_{1,n}p_{3,n}$	$p_{2,n}^2$	$2p_{2,n}p_{3,n}$	$p_{3,n}^2$
Mean phenotypic value	μ_{11}	μ_{12}	μ_{13}	μ_{22}	μ_{23}	μ_{33}

The mathematics is carried out with the same assumptions stated in the beginning of this section under the one locus with two alleles case.

It is seen in this case of three alleles that there are many possible configurations of the genotypic values μ_{ij} that each have to be considered because each leads to a different overall response relation of the population to the number of generations. In this case of three alleles, the number of different configurations is large. We have not enumerated all the possibilities, but we may note that there are $6!$ possible orders of the six values μ_{ij} . These will occur in sets of $3!$, because any permutation of labels of the genes will give the same mathematical structure. In addition there are many possibilities of dominance relations, as for instance that A_1 is completely dominant to A_2 so that $\mu_{11} = \mu_{12} > \mu_{22}$. At the same time we could have A_1 only partially dominant to A_3 , so that $\mu_{11} > \mu_{13} > \mu_{33}$. It would still be necessary to place μ_{23} in relation to μ_{11} , μ_{12} , μ_{13} and μ_{33} . We have found it necessary therefore to restrict ourselves to a few very simple situations. The ones which seemed amenable to treatment are examined in this section.

No dominance case with equal effects

The mean phenotypic values of the genotypes are presented in Table 7. The numerical values in the body of the table determine only the order of the genotypes. The mathematics that follows does not depend on these particular values.

Table 7. Mean phenotypic values of the genotypes.

	A_1	A_2	A_3
A_1	1	2	3
A_2	2	3	4
A_3	3	4	5

Accordingly, the order of the genotypes is $A_1A_1 < A_1A_2 < A_1A_3 = A_2A_2 < A_2A_3 < A_3A_3$. Following the approach of the one locus with two alleles we find five possible stages of progress.

$$1. \quad (1-S) < p_{1,n}^2.$$

The allelic frequencies in the next generation are given by

$$\begin{aligned}
 p_{3,n+1} &= \frac{1}{S} [p_{3,n}^2 + p_{2,n}p_{3,n} + p_{1,n}p_{3,n}] = \frac{1}{S} p_{3,n} \\
 p_{2,n+1} &= \frac{1}{S} [p_{2,n}p_{3,n} + p_{2,n}^2 + p_{1,n}p_{2,n}] = \frac{1}{S} p_{2,n} \\
 p_{1,n+1} &= \frac{1}{S} \{p_{1,n}p_{3,n} + p_{1,n}p_{2,n} + [p_{1,n}^2 - (1-S)]\} \\
 &= \frac{1}{S} [p_{1,n} - 1 + S] .
 \end{aligned} \tag{2.1}$$

So, as long as the inequality on S holds,

$$\begin{aligned}
 p_{3,n} &= \frac{1}{S^n} p_{3,o} \\
 p_{2,n} &= \frac{1}{S^n} p_{2,o} \\
 p_{1,n} &= \frac{1}{S^n} [p_{1,o} - (1-S^n)] ,
 \end{aligned} \tag{2.2}$$

where $p_{i,0}$ is the frequency of the A_i allele in the initial population. It is seen from (2.2) that when selection is strong, i.e. S is small, the change in the frequencies will be rapid and no equilibrium can be reached. The change in the allelic frequencies will continue until $(1-S) \geq p_{1,n}^2$. Using (2.2) and noting that $p_1 = 1 - p_2 - p_3$, we find

$$(1-S) \geq [1 - S^n(p_{2,0} + p_{3,0})]^2 \quad (2.3)$$

By taking the square root and then the logarithms of each side of (2.3) we find

$$n \ln S \geq \ln(1 - \sqrt{1-S}) - \ln(p_{2,0} + p_{3,0}).$$

The number of generations required for the first stage is given by

$$n \doteq \frac{1}{\ln S} [\ln(1 - \sqrt{1-S}) - \ln(p_{2,0} + p_{3,0})]. \quad (2.4)$$

The number will be small if S is small and it increases with S .

A linear relationship between the allelic frequencies and the generation number can be obtained by taking the logarithm of each side of (2.2)

$$\begin{aligned} \ln p_{3,n} &= \ln p_{3,0} - (\ln S) n \\ \ln p_{2,n} &= \ln p_{2,0} - (\ln S) n \\ \ln(1-p_{1,n}) &= \ln(1-p_{1,0}) - (\ln S) n. \end{aligned} \quad (2.5)$$

Equations (2.5) define functions of the allelic frequencies which change linearly with generations with slope given by

$$\hat{\beta} = - \ln S , \quad (2.6)$$

which is a function of selection.

$$2. \quad p_{1,n}^2 \leq (1-S) < p_{1,n}^2 + 2p_{1,n}p_{2,n} .$$

We find

$$\begin{aligned} p_{3,n+1} &= \frac{1}{S} [p_{3,n}^2 + p_{2,n}p_{3,n} + p_{1,n}p_{3,n}] = \frac{1}{S} p_{3,n} \\ p_{2,n+1} &= \frac{1}{S} \{p_{2,n}p_{3,n} + p_{2,n}^2 + \frac{1}{2} [2p_{1,n}p_{2,n} - (1-S - p_{1,n}^2)]\} \\ &= \frac{1}{S} [p_{2,n} - \frac{1}{2} (1-S - p_{1,n}^2)] \\ p_{1,n+1} &= \frac{1}{S} \{p_{1,n}p_{3,n} + \frac{1}{2} [2p_{1,n}p_{2,n} - (1-S - p_{1,n}^2)]\} \\ &= \frac{1}{2S} [p_{1,n}(2 - p_{1,n}) - (1-S)] . \end{aligned} \quad (2.7)$$

The change in the allelic frequencies is given by

$$\Delta p_{i,n} = p_{i,n+1} - p_{i,n} \quad i = 1, 2, 3 .$$

Using equations (2.7), we find

$$\begin{aligned} \Delta p_{3,n} &= \frac{1-S}{S} p_{3,n} \\ \Delta p_{1,n} &= \frac{1}{2S} [- (1-S) + 2(1-S)p_{1,n} - p_{1,n}^2] . \end{aligned} \quad (2.8)$$

The change in frequencies may be continuized by writing

$$\begin{aligned} \Delta p_{3,n} &= \frac{1-S}{S} p_{3,n} \doteq \frac{dp_{3,n}}{dn} \\ \Delta p_{1,n} &= \frac{1}{2S} [- (1-S) + 2(1-S)p_{1,n} - p_{1,n}^2] \doteq \frac{dp_{1,n}}{dn} . \end{aligned} \quad (2.9)$$

A solution of the differential equations (2.9) is given by

$$\ln p_{3,n} = \ln p_{3,o} + \frac{1-S}{S} n \quad (2.10)$$

$$\tan^{-1} \frac{1-S-p_{1,n}}{\sqrt{S(1-S)}} = \tan^{-1} \frac{1-S-p_{1,o}}{\sqrt{S(1-S)}} + \frac{1}{2} \sqrt{\frac{1-S}{S}} n. \quad (2.11)$$

Equations (2.10) and (2.11) define functions of the allelic frequencies p_3 and p_1 which are expected to change linearly with generations with slopes given by $\frac{1-S}{S}$ and $\frac{1}{2} \sqrt{\frac{1-S}{S}}$, respectively. If $(1-S) - \sqrt{S(1-S)} < p_1 < (1-S) + \sqrt{S(1-S)}$, equation (2.11) may be written in the approximate form

$$\frac{1-S-p_{1,n}}{\sqrt{S(1-S)}} = \frac{1-S-p_{1,o}}{\sqrt{S(1-S)}} + \frac{1}{2} \sqrt{\frac{1-S}{S}} n. \quad (2.12)$$

In the approximation given by (2.12), higher ordered terms in p_1 are ignored. It is seen from (2.7) and (2.10) that the change in frequencies will continue to the next stage.

$$3. \quad p_{1,n}^2 + 2p_{1,n}p_{2,n} \leq (1-S) < p_{1,n}^2 + 2p_{1,n}p_{2,n} + 2p_{1,n}p_{3,n} + p_{2,n}^2.$$

In this stage all of the A_1A_1 and A_1A_2 and some of the A_1A_3 and A_2A_2 individuals are culled. Thus there are A_1A_3 and A_2A_2 among the individuals saved. The array among culled individuals in generation n is

$$\frac{2p_{1,n}p_{3,n} A_1A_3 + p_{2,n}^2 A_2A_2}{2p_{1,n}p_{3,n} + p_{2,n}^2}.$$

The array among the saved individuals in generation n is

$$\frac{1}{S} \{ p_{3,n}^2 A_3 A_3 + 2p_{2,n} p_{3,n} A_2 A_3 + [p_{2,n}^2 - (1-S - p_{1,n}^2 - 2p_{1,n} p_{2,n})] \frac{p_{2,n}^2}{2p_{1,n} p_{3,n} + p_{2,n}^2} \} A_2 A_2 + [2p_{1,n} p_{3,n} - (1-S - p_{1,n}^2 - 2p_{1,n} p_{2,n})] \frac{2p_{1,n} p_{3,n}}{2p_{1,n} p_{3,n} + p_{2,n}^2} \} A_1 A_3 \} .$$

Consequently, we find

$$p_{3,n+1} = \frac{1}{S} \{ p_{3,n}^2 + p_{2,n} p_{3,n} + \frac{1}{2} [2p_{1,n} p_{3,n} - (1-S - p_{1,n}^2 - 2p_{1,n} p_{2,n})] \frac{2p_{1,n} p_{3,n}}{2p_{1,n} p_{3,n} + p_{2,n}^2} \}$$

$$= \frac{1}{S} (p_{3,n} - p_{1,n} p_{3,n} A)$$

$$p_{2,n+1} = \frac{1}{S} \{ p_{2,n} p_{3,n} + [p_{2,n}^2 - (1-S - p_{1,n}^2 - 2p_{1,n} p_{2,n})]$$

$$\frac{p_{2,n}^2}{2p_{1,n} p_{3,n} + p_{2,n}^2} \} \} \quad (2.13)$$

$$= \frac{1}{S} [p_{2,n} (1-p_{1,n}) - p_{2,n}^2 A]$$

$$p_{1,n+1} = \frac{1}{S} \{ p_{1,n} p_{3,n} - (1-S - p_{1,n}^2 - 2p_{1,n} p_{2,n}) \frac{p_{1,n} p_{3,n}}{2p_{1,n} p_{3,n} + p_{2,n}^2} \}$$

$$= \frac{1}{S} p_{1,n} p_{3,n} (1-A) ,$$

$$\text{where } A = \frac{1-S - p_{1,n}^2 - 2p_{1,n} p_{2,n}}{2p_{1,n} p_{3,n} + p_{2,n}^2} .$$

$$4. \quad 1 - 2p_{2,n}p_{3,n} - p_{3,n}^2 \leq (1-S) < 1 - p_{3,n}^2.$$

We find

$$\begin{aligned} p_{3,n+1} &= \frac{1}{S} \{ p_{3,n}^2 + \frac{1}{2} [2p_{2,n}p_{3,n} - (1-S - 1 + 2p_{2,n}p_{3,n} + p_{3,n}^2)] \} \\ &= \frac{1}{2} + \frac{1}{2S} p_{3,n}^2 \\ p_{2,n+1} &= \frac{1}{2S} (S - p_{3,n}^2) \\ p_{1,n+1} &= 0. \end{aligned} \tag{2.14}$$

It is seen from (2.14) that the frequency p_3 of A_3 will continue to increase and will approach unity as p_3 approaches \sqrt{S} .

The change in the allelic frequencies is given by

$$\Delta p_{i,n} = p_{i,n+1} - p_{i,n} \quad i = 2, 3.$$

Using (2.14) we find

$$\Delta p_{3,n} = \frac{1}{2S} [S - 2Sp_{3,n} + p_{3,n}^2]. \tag{2.15}$$

Equation (2.15) may be continuized by writing

$$\Delta p_{3,n} = \frac{1}{2S} [S - 2Sp_{3,n} + p_{3,n}^2] \doteq \frac{dp_{3,n}}{dn}. \tag{2.16}$$

The above equation (2.16) has the solution

$$\tan^{-1} \frac{p_{3,n} - S}{\sqrt{S(1-S)}} = \tan^{-1} \frac{p_{3,0} - S}{\sqrt{S(1-S)}} + \frac{1}{2} \sqrt{\frac{1-S}{S}} n. \tag{2.17}$$

Equation (2.17) defines a function of p_3 which is expected to change linearly with generations with slope given by $\frac{1}{2} \sqrt{\frac{1-S}{S}}$. if

$$S - \sqrt{S(1-S)} < p_3 < S + \sqrt{S(1-S)} ,$$

equation (2.17) may be written in the approximate form

$$\frac{p_{3,n} - S}{\sqrt{S(1-S)}} = \frac{p_{3,o} - S}{\sqrt{S(1-S)}} + \frac{1}{2} \sqrt{\frac{1-S}{S}} n , \quad (2.18)$$

or

$$p_{3,n} = p_{3,o} + \frac{1}{2} (1-S) n .$$

In the above approximation in (2.18), higher ordered terms in p_3 are ignored.

$$5. \quad (1-S) \geq 1 - p_{3,n}^2 .$$

After one generation we find that the population will consist solely of A_3A_3 individuals and the allelic frequencies at equilibrium are $p_3 = 1$ and $p_1 = p_2 = 0$.

Complete dominance with equally spaced effects

The mean phenotypic values of the genotypes are given in Table 8. The numerical values in the body of the table determine only the ordering of the genotypes. The mathematics that follows depends only on the ordering.

Table 8. Mean phenotypic values of genotypes .

	A ₁	A ₂	A ₃
A ₁	1	2	3
A ₂	2	2	3
A ₃	3	3	3

The order of the genotypes is $A_1A_1 < A_1A_2 = A_2A_2 < A_1A_3 = A_2A_3 = A_3A_3$.

Accordingly, we have three possible stages of progress.

$$1. \quad (1-S) < p_{1,n}^2.$$

We find

$$\begin{aligned}
 p_{3,n+1} &= \frac{1}{S} [p_{3,n}^2 + p_{2,n}p_{3,n} + p_{1,n}p_{3,n}] = \frac{1}{S} p_{3,n} \\
 p_{2,n+1} &= \frac{1}{S} [p_{2,n}p_{3,n} + p_{2,n}^2 + p_{1,n}p_{2,n}] = \frac{1}{S} p_{2,n} \\
 p_{1,n+1} &= \frac{1}{S} \{p_{1,n}p_{3,n} + p_{1,n}p_{2,n} + [p_{1,n}^2 - (1-S)]\} \\
 &= \frac{1}{S} [p_{1,n} - (1-S)] .
 \end{aligned} \tag{2.19}$$

Equations (2.19) are the same as (2.1) obtained under the first stage for the case of no dominance. Thus equations (2.2) to (2.6) apply also to this stage.

$$2. \quad p_{1,n}^2 \leq (1-S) < (p_{1,n} + p_{2,n})^2.$$

In this stage all of the A_1A_1 and some of the A_1A_2 and A_2A_2 individuals are culled. Thus there are A_1A_2 and A_2A_2 among the individuals saved. The array among the culled individuals in generation n is

$$\frac{2p_{1,n}p_{2,n} A_1 A_2 + p_{2,n}^2 A_2 A_2}{2p_{1,n}p_{2,n} + p_{2,n}^2} = \frac{2p_{1,n} A_1 A_2 + p_{2,n} A_2 A_2}{1 + p_{1,n} - p_{3,n}}$$

The array among the saved individuals in generation n is

$$\begin{aligned} & \frac{1}{S} \{ p_{3,n}^2 A_3 A_3 + 2p_{2,n}p_{3,n} A_2 A_3 + 2p_{1,n}p_{3,n} A_1 A_3 \\ & + [p_{2,n}^2 - (1-S - p_{1,n}^2) \frac{p_{2,n}}{1 + p_{1,n} - p_{3,n}}] A_2 A_2 \\ & + [2p_{1,n}p_{2,n} - (1-S - p_{1,n}^2) \frac{2p_{1,n}}{1 + p_{1,n} - p_{3,n}}] A_1 A_2 \cdot \end{aligned}$$

Thus we have

$$\begin{aligned} p_{3,n+1} &= \frac{1}{S} \{ p_{3,n}^2 + p_{2,n}p_{3,n} + p_{1,n}p_{3,n} \} \\ &= \frac{1}{S} p_{3,n} \\ p_{2,n+1} &= \frac{1}{S} \{ p_{2,n}p_{3,n} + p_{2,n}^2 - \frac{p_{2,n}(1-S - p_{1,n}^2)}{1 + p_{1,n} - p_{3,n}} \\ & \quad + p_{1,n}p_{2,n} - \frac{p_{1,n}(1-S - p_{1,n}^2)}{1 + p_{1,n} - p_{3,n}} \} \\ &= \frac{1}{S} \{ p_{2,n} - \frac{(p_{2,n} + p_{1,n})(1-S - p_{1,n}^2)}{1 + p_{1,n} - p_{3,n}} \} \quad (2.20) \\ p_{1,n+1} &= \frac{1}{S} \{ p_{1,n}p_{3,n} + p_{1,n}p_{2,n} - \frac{p_{1,n}(1-S - p_{1,n}^2)}{1 + p_{1,n} - p_{3,n}} \} \\ &= \frac{1}{S} [p_{1,n}(1-p_{1,n}) - \frac{p_{1,n}(1-S - p_{1,n}^2)}{1 + p_{1,n} - p_{3,n}}] \cdot \end{aligned}$$

The change in the allelic frequencies is given by

$$\Delta p_{i,n} = p_{i,n+1} - p_{i,n} \quad i = 1, 2, 3.$$

Using (2.20), we find

$$\begin{aligned} \Delta p_{3,n} &= \frac{1-S}{S} p_{3,n} \\ \Delta p_{2,n} &= \frac{1}{S(1 + p_{1,n} - p_{3,n})} \{ p_{2,n}(1-S)(1 + p_{1,n} - p_{3,n}) \\ &\quad - (1-p_{3,n})(1-S - p_{1,n}^2) \} \end{aligned} \quad (2.21)$$

$$\Delta p_{1,n} = \frac{1}{S(1 + p_{1,n} - p_{3,n})} p_{1,n} [S(p_{3,n} - p_{1,n}) - p_{3,n}(1-p_{1,n})].$$

It is seen from (2.20) and (2.21) that the change in the frequencies will continue to the next stage.

$$3. \quad (1-S) \geq (p_{1,n} + p_{2,n})^2.$$

We find

$$\begin{aligned} p_{3,n+1} &= \frac{1}{W} [p_{3,n}^2 + p_{2,n}p_{3,n} + p_{1,n}p_{3,n}] \\ &= \frac{1}{W} p_{3,n} \\ p_{2,n+1} &= \frac{1}{W} [p_{2,n}p_{3,n}] \\ p_{1,n+1} &= \frac{1}{W} p_{1,n}p_{3,n}, \end{aligned} \quad (2.22)$$

where

$$\begin{aligned}
W &= 1 - (p_{1,n} + p_{2,n})^2 \\
&= 2p_{1,n}p_{3,n} + 2p_{2,n}p_{3,n} + p_{3,n}^2 \\
&= p_{3,n}(2p_{1,n} + 2p_{2,n} + p_{3,n}) \\
&= p_{3,n}(1 + p_{1,n} + p_{2,n}) .
\end{aligned}$$

It is seen from (2.22) that the ratio

$$\frac{p_{2,n}}{p_{1,n}} = \frac{p_{2,0}}{p_{1,0}}$$

will remain constant over generations. Consider now the ratio

$$\frac{p_{2,n+1}}{p_{3,n+1}} = p_{2,n} .$$

Thus

$$\frac{p_{2,1}}{p_{3,1}} = p_{2,0}$$

$$\frac{p_{2,2}}{p_{3,2}} = p_{2,1} = p_{2,0} p_{3,1}$$

$$\frac{p_{2,3}}{p_{3,3}} = p_{2,2} = p_{2,0} p_{3,1} p_{3,2} .$$

In general we have

$$p_{2,n} = p_{2,0} p_{3,1} p_{3,2} p_{3,3} \cdots p_{3,n} . \quad (2.23)$$

Since $0 < p < 1$, we find that $p_{2,n}$ will approach zero slowly as n increases. Equations (2.22) may be written in the form

$$\begin{aligned} p_{3,n+1} &= \frac{1}{1 + p_{1,n} + p_{2,n}} = \frac{1}{2 - p_{3,n}} \\ p_{2,n+1} &= \frac{p_{2,n}}{1 + p_{1,n} + p_{2,n}} \\ p_{1,n+1} &= \frac{p_{1,n}}{1 + p_{1,n} + p_{2,n}} . \end{aligned} \quad (2.24)$$

The first equation for p_3 in (2.24) may be written in the form

$$p_{3,n} = \frac{p_{3,0} + n(1-p_{3,0})}{1 + n(1-p_{3,0})} = \frac{n - (n-1)p_{3,0}}{(n+1) - np_{3,0}} . \quad (2.25)$$

It is seen from (2.25) that p_3 will approach unity as n gets large. Writing the second equation for p_2 in (2.24) in the form

$$p_{2,n+1} = \frac{p_{2,n}}{2 - p_{3,n}}$$

and using (2.25), we find

$$p_{2,n} = A_1 A_2 \cdots A_n p_{2,0} , \quad (2.26)$$

where

$$A_i = \frac{1 + i(1-p_{3,0})}{2 + i(1-p_{3,0}) - p_{3,0}} .$$

Equations (2.25) and (2.26) define p_3 and p_2 in terms of n and the initial frequency $p_{3,0}$. It is seen from (2.26) that A_1 approaches zero as i , the generation number, gets large. Consequently, p_2 will approach zero as the number of generations gets larger.

Overdominance

Two models are considered. The mean phenotypic values for the two models are given in Table 9.

Table 9. Mean phenotypic values of genotypes for the overdominance model.

	Model I			Model II		
	A_1	A_2	A_3	A_1	A_2	A_3
A_1	1	4	5	1	2	2
A_2	4	2	6	2	1	2
A_3	5	6	3	2	2	1

Model I: Overdominance with additive effect. We consider first

Model I. The order of the genotypes may be written as $A_1A_1 < A_2A_2 < A_3A_3 < A_1A_2 < A_1A_3 < A_2A_3$. Accordingly, there are six possible stages of progress.

1. $(1-S) < p_{1,n}^2$.

This stage is the same as the first stage discussed above for the intermediate dominance. Accordingly equations (2.1) to (2.6) as well as the discussion apply to this stage.

$$2. \quad p_{1,n}^2 \leq (1-S) < p_{1,n}^2 + p_{2,n}^2.$$

After one generation we find

$$\begin{aligned} p_{3,n+1} &= \frac{1}{S} [p_{3,n}^2 + p_{1,n}p_{3,n} + p_{2,n}p_{3,n}] = \frac{p_{3,n}}{S} \\ p_{2,n+1} &= \frac{1}{S} \{p_{2,n}p_{3,n} + p_{1,n}p_{2,n} + [p_{2,n}^2 - (1-S - p_{1,n}^2)]\} \\ &= \frac{1}{S} [p_{2,n} - (1-S) + p_{1,n}^2] \\ p_{1,n+1} &= \frac{1}{S} [p_{1,n}p_{3,n} + p_{1,n}p_{2,n}] = \frac{1}{S} p_{1,n}(1-p_{1,n}). \end{aligned} \quad (2.27)$$

So, as long as the inequality on S holds, we have

$$p_{3,n} = S^{-n} p_{3,0}. \quad (2.28)$$

$$\ln \left| \frac{p_{1,n}}{1-S - p_{1,n}} \right| = \ln \left| \frac{p_{1,0}}{1-S - p_{1,0}} \right| + \frac{1-S}{S} n. \quad (2.29)$$

It is seen from (2.27) and (2.28) that the frequencies will continue to change and no equilibrium can be reached in this stage.

$$3. \quad p_{1,n}^2 + p_{2,n}^2 \leq (1-S) < p_{1,n}^2 + p_{2,n}^2 + p_{3,n}^2.$$

We find

$$\begin{aligned} p_{3,n+1} &= \frac{1}{S} \{p_{2,n}p_{3,n} + p_{1,n}p_{3,n} + [p_{3,n}^2 - (1-S - p_{1,n}^2 - p_{2,n}^2)]\} \\ &= \frac{1}{S} [p_{3,n} - (1-S) + p_{1,n}^2 + p_{2,n}^2] \\ p_{2,n+1} &= \frac{1}{S} [p_{2,n}p_{3,n} + p_{1,n}p_{2,n}] = \frac{1}{S} p_{2,n}(1-p_{2,n}) \\ p_{1,n+1} &= \frac{1}{S} [p_{1,n}p_{2,n} + p_{1,n}p_{3,n}] = \frac{1}{S} p_{1,n}(1-p_{1,n}). \end{aligned} \quad (2.30)$$

Equilibrium in this stage exists when $\Delta p_{i,n} = p_{i,n+1} - p_{i,n} = 0$ for $i = 1, 2, 3$. Using (2.30), we find

$$1 - p_1 - S = 0 \quad (2.31)$$

$$1 - p_2 - S = 0$$

$$p_3(1-S) - (1-S) + p_1^2 + p_2^2 = 0 .$$

Equations (2.31) will be satisfied for the non-trivial equilibrium values $p_1 = p_2 = 1-S$ and $p_3 = 2S-1$. If there is equilibrium in stage three, we see from the defining inequalities for the stage and from (2.31) that S must be greater than $2/3$. Table 10 gives the equilibrium frequencies for some values of S greater than $2/3$.

Table 10. Equilibrium values for different degrees of selection (S).

	S				
	.667	.7	.75	.8	.9
p_1	.333	.3	.25	.2	.1
p_2	.333	.3	.25	.2	.1
p_3	.334	.4	.50	.6	.8

$$4. \quad p_{1,n}^2 + p_{2,n}^2 + p_{3,n}^2 \leq (1-S) < 1 - 2p_{1,n}p_{3,n} - 2p_{2,n}p_{3,n}.$$

After one generation we find

$$p_{3,n+1} = \frac{1}{S} (p_{1,n}p_{3,n} + p_{2,n}p_{3,n}) = \frac{1}{S} p_{3,n}(1-p_{3,n})$$

$$p_{2,n+1} = \frac{1}{S} \{p_{2,n}p_{3,n} + \frac{1}{2} [2p_{1,n}p_{2,n} - (1-S - p_{1,n}^2 - p_{2,n}^2 - p_{3,n}^2)]\}$$

$$= \frac{1}{2S} [p_{2,n}(2-p_{2,n}) + (p_{1,n}^2 + p_{3,n}^2) - (1-S)] \quad (2.32)$$

$$\begin{aligned} p_{1,n+1} &= \frac{1}{S} \{p_{1,n}p_{3,n} + \frac{1}{2} [2p_{1,n}p_{2,n} - (1-S - p_{1,n}^2 - p_{2,n}^2 - p_{3,n}^2)]\} \\ &= \frac{1}{2S} [p_{1,n}(2-p_{1,n}) + (p_{2,n}^2 + p_{3,n}^2) - (1-S)] . \end{aligned}$$

Equilibrium in this stage exists when $\Delta p_{i,n} = p_{i,n+1} - p_{i,n} = 0$ for all i .

Using (2.32), the condition requires

$$1 - p_3 - S = 0$$

$$p_2(2-2S-p_2) + p_1^2 + p_3^2 - (1-S) = 0 \quad (2.33)$$

$$p_1(2-2S-p_1) + p_2^2 + p_3^2 - (1-S) = 0 .$$

It is seen from (2.33) the non-trivial equilibrium values are given by

$p_3 = 1-S$ and $p_1 = p_2 = \frac{S}{2}$. If there is equilibrium in stage four, we see

from the defining inequalities for the stage and from (2.33) that

$\frac{1}{2} < S \leq \frac{2}{3}$. Table 11 gives the equilibrium values for different values of S .

Table 11. Equilibrium values for different values of selection (S).

	S			
	.5	.55	.6	.667
p_1	.25	.225	.3	.333
p_2	.25	.225	.3	.333
p_3	.50	.450	.4	.333

$$5. \quad 1 - 2p_{1,n}p_{3,n} - 2p_{2,n}p_{3,n} \leq (1-S) < 1 - 2p_{2,n}p_{3,n}.$$

We find

$$\begin{aligned} p_{3,n+1} &= \frac{1}{S} \{p_{2,n}p_{3,n} + \frac{1}{2} [2p_{1,n}p_{3,n} - (1-S - 1 + 2p_{1,n}p_{3,n} + wp_{2,n}p_{3,n})]\} \\ &= \frac{1}{2} \end{aligned}$$

$$p_{2,n+1} = \frac{1}{S} p_{2,n}p_{3,n} \tag{2.34}$$

$$\begin{aligned} p_{1,n+1} &= \frac{1}{2S} \{2p_{1,n}p_{3,n} - [(1-S) - (1-2p_{1,n}p_{3,n} - 2p_{2,n}p_{3,n})]\} \\ &= \frac{1}{2S} [S - 2p_{2,n}p_{3,n}]. \end{aligned}$$

In the next generation we find

$$\begin{aligned} p_{3,n+2} &= \frac{1}{2} \\ p_{2,n+2} &= \frac{1}{2S} p_{2,n} \\ p_{1,n+2} &= \frac{1}{2S} (S - p_{2,n}). \end{aligned} \tag{2.35}$$

It is seen from (2.34) and (2.35) that the frequency p_3 of the A_3 allele approaches its equilibrium value of $\frac{1}{2}$ in this stage and that p_2 will continue to increase up to the next stage.

$$6. \quad (1-S) \geq 1 - 2p_{2,n}p_{3,n}.$$

After one generation the population will consist solely of A_2A_3 individuals. Thus we have at equilibrium $p_2 = p_3 = \frac{1}{2}$ and $p_1 = 0$. If there is equilibrium in stage six, we see from the defining inequality for the stage that $0 < S \leq \frac{1}{2}$.

Table 12 sums up the equilibrium points for the allelic frequencies for different degrees of selection.

Table 12. Equilibrium points corresponding to proportions saved (S).

	$0 < S \leq \frac{1}{2}$	$\frac{1}{2} < S \leq \frac{2}{3}$	$\frac{2}{3} < S < 1$
p_1	0	$\frac{S}{2}$	$1-S$
p_2	$\frac{1}{2}$	$\frac{S}{2}$	$1-S$
p_3	$\frac{1}{2}$	$1-S$	$2S-1$

For the case of one locus with two alleles, Pollak (1966) showed that, when overdominance is present, it is possible for the population mean to actually decline during selection, in spite of the fact that in each generation only individuals having high genotypic values are chosen to reproduce. Using the mean phenotypic values given in Table 6, the mean of the population in generation n is given by

$$W = p_{1,n}^2 \mu_{11} + 2p_{1,n}p_{2,n} \mu_{12} + 2p_{1,n}p_{3,n} \mu_{13} \\ + p_{2,n}^2 \mu_{22} + 2p_{2,n}p_{3,n} \mu_{23} + p_{3,n}^2 \mu_{33} .$$

The development of the present study given above shows that Pollak's findings do hold for the one locus with three alleles case. This is illustrated by the following example, in which the mean phenotypic values are given in Table 9--Model I and $S = .8$. The calculations are summarized in Table 13.

Table 13. The change in the allelic frequencies (p_i) and the population mean (W) with generations (n).

n	0	1	2	3	4
p_1	.0769	.0887	.1012	.1137	.1293
p_2	.3846	.2959	.2604	.2407	.2203
p_3	.5385	.6154	.6384	.6456	.6504
W	4.31	4.26	4.22	4.20	4.19

It is seen from Table 13 that the population is initially in its optimum condition and deteriorates thereafter. Thus Pollak's results for the simple one locus with two alleles as well as the present findings for the three alleles case differ from those that would be expected from the classical theory discussed by Fisher (1930), in which a random mating population has maximum mean value when the gene frequencies are at stable equilibrium values.

Model II: Complete overdominance with equal effects. We now consider the mean phenotypic values of the genotypes given by Model II, Table 9. The order of the genotypes is $A_1A_1 = A_2A_2 = A_3A_3 < A_1A_2 = A_1A_3 = A_2A_3$. There are two possible stages of progress.

$$1. \quad (1-S) < p_{1,n}^2 + p_{2,n}^2 + p_{3,n}^2.$$

In this case some of the A_1A_1 , A_2A_2 and A_3A_3 are culled. The array among the culled individuals in generation n is

$$\frac{p_{1,n}^2 A_1A_1 + p_{2,n}^2 A_2A_2 + p_{3,n}^2 A_3A_3}{p_{1,n}^2 + p_{2,n}^2 + p_{3,n}^2}.$$

The array among the saved individuals in generation n is

$$\begin{aligned}
 & \frac{1}{S} \{ 2p_{2,n}p_{3,n} A_2 A_3 + 2p_{1,n}p_{3,n} A_1 A_3 + 2p_{1,n}p_{2,n} A_1 A_2 \\
 & + (p_{3,n}^2 - \frac{(1-S) p_{3,n}^2}{p_{1,n}^2 + p_{2,n}^2 + p_{3,n}^2}) A_3 A_3 \\
 & + (p_{2,n}^2 - \frac{(1-S) p_{2,n}^2}{p_{1,n}^2 + p_{2,n}^2 + p_{3,n}^2}) A_2 A_2 \\
 & + (p_{1,n}^2 - \frac{(1-S) p_{1,n}^2}{p_{1,n}^2 + p_{2,n}^2 + p_{3,n}^2}) A_1 A_1 \} .
 \end{aligned}$$

Thus we find

$$\begin{aligned}
 p_{3,n+1} &= \frac{1}{S} [p_{2,n}p_{3,n} + p_{1,n}p_{3,n} + p_{3,n}^2 - (1-S)p_{3,n}^2 W^{-1}] \\
 &= \frac{1}{S} [p_{3,n} - (1-S)p_{3,n}^2 W^{-1}] \\
 p_{2,n+1} &= \frac{1}{S} [p_{2,n}p_{3,n} + p_{1,n}p_{2,n} + p_{2,n}^2 - (1-S)p_{2,n}^2 W^{-1}] \\
 &= \frac{1}{S} [p_{2,n} - (1-S)p_{2,n}^2 W^{-1}] \\
 p_{1,n+1} &= \frac{1}{S} [p_{1,n}p_{3,n} + p_{1,n}p_{2,n} + p_{1,n}^2 - (1-S)p_{1,n}^2 W^{-1}] \\
 &= \frac{1}{S} [p_{1,n} - (1-S)p_{1,n}^2 W^{-1}] ,
 \end{aligned} \tag{2.36}$$

where

$$W = p_{1,n}^2 + p_{2,n}^2 + p_{3,n}^2 .$$

The change in the allelic frequencies is given by

$$\Delta p_{i,n} = p_{i,n+1} - p_{i,n} \quad i = 1, 2, 3.$$

Using (2.36), we find

$$\Delta p_{i,n} = \frac{1-S}{S} p_{i,n} \left[1 - \frac{p_{i,n}}{p_{1,n}^2 + p_{2,n}^2 + p_{3,n}^2} \right]. \quad (2.37)$$

Equilibrium exists when $\Delta p_{i,n} = 0$ for all i . It is seen from (2.37) that equilibrium exists when $p_1 = p_2 = p_3 = \frac{1}{3}$.

$$2. \quad (1-S) \geq p_{1,n}^2 + p_{2,n}^2 + p_{3,n}^2.$$

We find

$$\begin{aligned} p_{3,n+1} &= \frac{1}{W} (p_{2,n}p_{3,n} + p_{1,n}p_{3,n}) = \frac{1}{W} p_{3,n}(1-p_{3,n}) \\ p_{2,n+1} &= \frac{1}{W} (p_{2,n}p_{3,n} + p_{1,n}p_{2,n}) = \frac{1}{W} p_{2,n}(1-p_{2,n}) \\ p_{1,n+1} &= \frac{1}{W} (p_{1,n}p_{3,n} + p_{1,n}p_{2,n}) = \frac{1}{W} p_{1,n}(1-p_{1,n}), \end{aligned} \quad (2.38)$$

where

$$W = 2(p_{1,n}p_{2,n} + p_{1,n}p_{3,n} + p_{2,n}p_{3,n}).$$

The change in the frequencies is given by

$$\Delta p_{i,n} = p_{i,n+1} - p_{i,n} \quad i = 1, 2, 3.$$

Using (2.38), we find

$$\Delta p_{i,n} = \frac{1}{W} p_{i,n} (1 - p_{i,n} - W) \quad i = 1, 2, 3. \quad (2.39)$$

Equilibrium exists when $\Delta p_{i,n} = 0$ for all i . It is seen from (2.39) that equilibrium exists when

$$1 - p_3 - W = 0$$

$$1 - p_2 - W = 0$$

$$1 - p_1 - W = 0 .$$

Thus equilibrium will be satisfied for $p_1 = p_2 = p_3 = \frac{1}{3}$ and it is stable.

Two-Loci Case

In the previous sections the effects of selection by culling on the change in gene frequency, the equilibrium conditions, and the approach to the equilibrium are examined for the one locus case. The study of one locus takes into account only inter-allelic effects such as additivity and dominance. In natural populations, however, natural selection is likely to occur for traits under the control of two or more loci. With two loci, two new factors add great complexity. They are linkage and interactions between loci. Such interactions between loci have been shown to exist in experimental studies (e.g. Dobzhansky and Spassky, 1967).

Some consequences of linkage in populations subject to selection were given by Fisher (1930). The effect of natural selection on gene frequencies for the two locus case has been studied by Kimura (1956) for the continuous time model and by Lewontin and Kojima (1960) for the discrete generation case. These workers assumed constant selective advantage of the genotypes with no environmental variability and no fecundity differences. The results of their investigations were sufficient to show that even for relatively

simple cases (two loci and symmetrical selective values) linkage may have profound effects on the course of natural selection.

The case of two loci under truncation selection with no environmental variability will be examined. In this case the genotypes are distinguishable with respect to the character. Following the one locus approach, we assume here that mating among the selected individuals is at random, and that the populations are infinite in size so that sampling element can be ignored. We assume that there are no differences in the fecundity of the matings. The fitness of an individual (measured in terms of selective advantage or probability that an individual with a given genotype will survive selection) is dependent on genotype at two linked loci A and B at each of which there are two alleles A,a and B,b. Selection applied is by truncation, with the best $S(0 < S < 1)$ individuals of each sex saved in each generation to produce the next generation through random mating. Whether or not an individual is culled would partially depend on his genotypic constitution relative to others at the time of selection, so that his "selective advantage" is frequency dependent. It is assumed that the ordering of phenotypic classes is the same in the two sexes. This type of selection is different from that considered by Lewontin and Kojima (1960), Parsons (1963) and Moran (1967). They examined the effect of natural selection using constant fitnesses (defined in terms of relative viability or the survival of zygotes to adults with no difference in fecundity) of genotypes throughout the selection period. Linkage between the two loci is assumed to be the same in the two sexes.

With two loci segregating there are four possible gametic types, AB, Ab, aB and ab with frequencies denoted by x_1 , x_2 , x_3 and x_4 , respectively. The gametic frequencies satisfy the following relations:

$$\begin{array}{ll}
 x_1 + x_2 = p & \text{the frequency of allele A} \\
 x_3 + x_4 = (1-p) & \text{the frequency of allele a} \\
 x_1 + x_3 = q & \text{the frequency of allele B} \\
 x_2 + x_4 = (1-q) & \text{the frequency of allele b} \\
 \sum x_i = 1 .
 \end{array}$$

Gametic disequilibrium, D, is defined usually by the difference between the product of the frequencies of the coupling gametes (AB,ab) and the product of the frequencies of the repulsion gametes (Ab,aB), i.e.

$$D = x_1 x_4 - x_2 x_3 .$$

Gametic equilibrium thus indicates that the alleles A and B are randomly associated in the population. Let x'_1 , x'_2 , x'_3 and x'_4 be the frequencies of the four gametes at gametic equilibrium such that

$$x'_1 x'_4 - x'_2 x'_3 = 0 .$$

Then we may write

$$\begin{array}{ll}
 x_1 = x'_1 + D & x_3 = x'_3 - D \\
 x_2 = x'_2 - D & x_4 = x'_4 + D .
 \end{array}$$

If we denote the recombination frequency between the two loci by r , then the frequencies of the ten possible genotypes and the gametes produced in generation n can be written as in Table 14.

Table 14. The frequencies of the genotypes and the gametes produced.

Genotypes	Genotypic frequencies	Gametic frequencies			
		AB	Ab	aB	ab
AB/AB	$x_{1,n}^2$	1	0	0	0
AB/Ab	$2x_{1,n}x_{2,n}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0
AB/aB	$2x_{1,n}x_{3,n}$	$\frac{1}{2}$	0	$\frac{1}{2}$	0
AB/ab	$2x_{1,n}x_{4,n}$	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$
Ab/aB	$2x_{2,n}x_{3,n}$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$
Ab/Ab	$x_{2,n}^2$	0	1	0	0
Ab/ab	$2x_{2,n}x_{4,n}$	0	$\frac{1}{2}$	0	$\frac{1}{2}$
aB/aB	$x_{3,n}^2$	0	0	1	0
aB/ab	$2x_{3,n}x_{4,n}$	0	0	$\frac{1}{2}$	$\frac{1}{2}$
ab/ab	$x_{4,n}^2$	0	0	0	1

For each model to be considered, the selection experiment was programmed on the computer and the gametic frequencies as well as gametic disequilibrium were obtained in every generation with no approximations used. Hence the numerical solutions obtained on the computer are considered to be exact and, therefore, are called exact numerical solutions throughout the discussion. Thus a comparison between the results obtained from the derived mathematical equations and the exact or computer results can be made to test the accuracy of the equations as well as the direction of the bias when it exists. For the numerical results the initial gametic frequencies $x_{2,0}$ and $x_{3,0}$ were taken to be equal. This would imply that $x_2 = x_3$ in

every generation of selection. The mean phenotypic values for the genotypes are presented in Table 27, Appendix B, for the genetic models considered.

With just two alleles at each locus, there are ten genotypic classes. If each of these has a specific phenotype different from every other phenotype there is a very large number of possible configurations. The work of this section is confined to two cases, one in which heterozygotes are assumed to be superior over both homozygotes and the other in which intermediate and complete dominance hold for both loci. For all genetic models considered two cases will be examined, one in which there are equal effects at the two loci, and the second in which the two loci are assumed to have unequal effects.

Two loci with equal effects

In this section the two loci are assumed to have equal effects with their initial frequency equal or unequal.

No dominance at both loci. The mean phenotypic values for the genotypes are presented in Table 27, Appendix B. The numerical values in the body of the table determine only the order of the genotypes. The mathematics that follows does not depend on the particular values used. Under this model the order of the genotypes is

$$aabb < Aabb = aaBb < AAbb = aaBB = AB/ab = Ab/aB < AABb = AaBB < AABB.$$

This case can be considered as arising from equal arithmetic genetic effects at the two loci. There are five possible stages of progress.

$$1. \quad (1-S) < x_{4,n}^2.$$

The genotypic frequencies in the next generation are given by

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} \{x_{1,n}^2 + x_{1,n} x_{2,n} + x_{1,n} x_{3,n} \\ &\quad + (1-r) x_{1,n} x_{4,n} + r x_{2,n} x_{3,n}\} \\ &= \frac{1}{S} \{x_{1,n} - r D_n\} \\ x_{2,n+1} &= \frac{1}{S} \{x_{1,n} x_{2,n} + r x_{1,n} x_{4,n} \\ &\quad + (1-r) x_{2,n} x_{3,n} + x_{2,n}^2 + x_{2,n} x_{4,n}\} \\ &= \frac{1}{S} \{x_{2,n} + r D_n\} \\ x_{3,n+1} &= \frac{1}{S} \{x_{1,n} x_{3,n} + r x_{1,n} x_{4,n} \\ &\quad + (1-r) x_{2,n} x_{3,n} + x_{3,n}^2 + x_{3,n} x_{4,n}\} \\ &= \frac{1}{S} \{x_{3,n} + r D_n\} \\ x_{4,n+1} &= \frac{1}{S} \{(1-r) x_{1,n} x_{4,n} + r x_{2,n} x_{3,n} + x_{2,n} x_{4,n} \\ &\quad + x_{3,n} x_{4,n} + x_{4,n}^2 - (1-S)\} \quad (3.1) \\ &= \frac{1}{S} \{x_{4,n} - (1-S) - r D_n\}. \end{aligned}$$

The gametic disequilibrium in the (n+1)th generation is given by

$$D_{n+1} = x_{1,n+1} x_{4,n+1} - x_{2,n+1} x_{3,n+1}.$$

Using equation (3.1), we find

$$D_{n+1} = S^{-2} \{D_n(1-rS) - (1-S) x_{1,n}\}. \quad (3.2)$$

In this stage an exact solution can be obtained by writing the equations for $x_{1,n+1}$, $x_{2,n+1}$, $x_{3,n+1}$ in (3.1) and D_{n+1} in (3.2) in the matrix form

$$\begin{bmatrix} x_{1,n+1} \\ x_{2,n+1} \\ x_{3,n+1} \\ D_{n+1} \end{bmatrix} = \begin{bmatrix} \frac{1}{S} & 0 & 0 & -\frac{r}{S} \\ 0 & \frac{1}{S} & 0 & \frac{r}{S} \\ 0 & 0 & \frac{1}{S} & \frac{r}{S} \\ -\frac{(1-S)}{S^2} & 0 & 0 & \frac{1-rS}{S^2} \end{bmatrix} \begin{bmatrix} x_{1,n} \\ x_{2,n} \\ x_{3,n} \\ D_n \end{bmatrix} \quad (3.3)$$

The frequencies x_1 , x_2 , x_3 and the gametic disequilibrium D in the second generation relate to those in the first generation by the equation

$$X_1 = A X_0, \quad (3.4)$$

where

$$X_1 = \begin{bmatrix} x_{1,1} \\ x_{2,1} \\ x_{3,1} \\ D_1 \end{bmatrix}, \quad X_0 = \begin{bmatrix} x_{1,0} \\ x_{2,0} \\ x_{3,0} \\ D_0 \end{bmatrix}, \quad A = \begin{bmatrix} \frac{1}{S} & 0 & 0 & \frac{r}{S} \\ 0 & \frac{1}{S} & 0 & \frac{r}{S} \\ 0 & 0 & \frac{1}{S} & \frac{r}{S} \\ -\frac{1-S}{S^2} & 0 & 0 & \frac{1-rS}{S^2} \end{bmatrix}$$

Using generation matrix methodology, it can be shown that

$$X_n = C D^n C^{-1} X_0, \quad (3.5)$$

where D is a diagonal matrix whose elements along the diagonal are the latent roots of the A matrix. The matrix C is any matrix satisfying the equation $C^{-1} A C = D$. The latent roots are given by

$$\frac{1}{S}, \frac{1}{S}, \frac{1}{S^2}, (1-r)/S.$$

The frequencies x_1, x_2, x_3 and the gametic disequilibrium D in generation n are given by

$$\begin{aligned} x_{1,n} &= \frac{1}{1-S+rS} \left[rS \left(\frac{1}{S^2} \right)^n + (1-S) \left(\frac{1-r}{S} \right)^n \right] x_{1,o} \\ &\quad - \frac{rS}{1-S+rS} \left[\left(\frac{1}{S^2} \right)^n - \left(\frac{1-r}{S} \right)^n \right] D_o \\ x_{2,n} &= \frac{(1-S) \left[1 - (1-r)^n \right] + rS(1-S^{-n})}{(1-S+rS) S^n} x_{1,o} \\ &\quad + \frac{1}{S^n} x_{2,o} + \frac{rS}{1-S+rS} \left[\left(\frac{1}{S^2} \right)^n - \left(\frac{1-r}{S} \right)^n \right] D_o \\ x_{3,n} &= \frac{(1-S) \left[1 - (1-r)^n \right] + rS(1-S^{-n})}{(1-S+rS) S^n} x_{1,o} \\ &\quad + \frac{1}{S^n} x_{3,o} + \frac{rS}{1-S+rS} \left[\left(\frac{1}{S^2} \right)^n - \left(\frac{1-r}{S} \right)^n \right] D_o \\ D_n &= - \frac{1-S}{1-S+rS} \left[\left(\frac{1}{S^2} \right)^n - \left(\frac{1-r}{S} \right)^n \right] x_{1,o} \\ &\quad + \frac{1}{1-S+rS} \left[(1-S) \left(\frac{1}{S^2} \right) + rS \left(\frac{1-r}{S} \right)^n \right] D_o \end{aligned} \tag{3.6}$$

The two roots $\frac{1}{S^2}$ and $\frac{(1-r)}{S}$ appearing in (3.6) show the relative importance of selection and linkage. For all the possible values of $S(0 < S < 1)$ and $r(0 < r < \frac{1}{2})$ we have

$$\frac{1}{S^2} > \frac{(1-r)}{S} \quad \text{or} \quad \frac{1}{S} > (1-r) .$$

The root $(1-r)/S$ which depends on linkage is very important since it is not possible for this root to be zero. When $r=0$, equations (3.6) reduce to

$$\begin{aligned} x_{1,n} &= \frac{1}{S^n} x_{1,0} \\ x_{2,n} &= \frac{1}{S^n} x_{2,0} \\ x_{3,n} &= \frac{1}{S^n} x_{3,0} \\ D_n &= - \left[\left(\frac{1}{S^2}\right)^n - \left(\frac{1}{S}\right)^n \right] x_{1,0} + \left(\frac{1}{S^2}\right)^n D_0 . \end{aligned} \tag{3.7}$$

For the case when linkage is complete, i.e. $r=0$, we find from (3.7)

$$\ln x_{i,n} = \ln x_{i,0} - (\ln S) n \quad i = 1, 2, 3 . \tag{3.8}$$

Equations (3.8) indicate that, when $r=0$, the function of the gametic frequency $\ln x_i (i = 1, 2, 3)$ is expected to change linearly with generations with slope given by $\hat{\beta}_1 = -\ln S$. Figure 4 shows the exact change in the function $\ln x_1$, with generations for different values of S and r when $D_0 = 0$. Table 29, Appendix B, shows the values for the exact slopes β , obtained from Figure 4, and their estimates $\hat{\beta}_1$. It is seen that the estimates are more accurate when r is small and S is large. The equation for

the gametic disequilibrium given in (3.5) shows that the selection may build up gametic disequilibrium under the additive model even when the initial population is in linkage equilibrium, i.e. $D_0 = 0$, and no linkage, i.e. $r = 1/2$. When either $D_0 = 0$ or $r = 0$, it is seen from (3.6) and (3.7) that selection is likely to build up negative gametic disequilibrium. Figure 3 shows the exact change in D_n under the additive model for different values of S and r when $D_0 = 0$. The values in the figure are the exact numerical solutions which are in complete agreement with those given by equation (3.6).

The frequency of the A and B genes in the n -th generation are given by

$$p_n = x_{1,n} + x_{2,n}$$

$$q_n = x_{1,n} + x_{3,n} ,$$

respectively. Using equation (3.6) we find

$$p_n = \frac{1}{s^n} (x_{1,0} + x_{2,0}) = \frac{1}{s^n} p_0$$

$$q_n = \frac{1}{s^n} (x_{1,0} + x_{3,0}) = \frac{1}{s^n} q_0 .$$
(3.9)

Equations (3.9) show that gene frequencies during the first stage change independently of linkage. The change in gene frequency given by (3.9) is the same as that obtained under the one locus case (equation 1.2). We now consider the difference

$$\begin{aligned}
q_n - p_n &= (x_{1,n} + x_{3,n}) - (x_{1,n} + x_{2,n}) \\
&= x_{3,n} - x_{2,n} \\
&= \frac{1}{s^n} (x_{3,o} - x_{2,o}) \\
&= \frac{1}{s^n} (q_o - p_o) .
\end{aligned}$$

The above equation shows that when the initial frequencies $x_{2,o} = x_{3,o}$ or $p_o = q_o$, the frequencies will remain equal in every generation during the first stage. When $x_{2,o} \neq x_{3,o}$ the difference $x_{3,n} - x_{2,n}$ will be increased by selection. With the initial population in Hardy-Weinberg equilibrium, the initial frequencies of the gametes satisfy the following relations

$$\begin{aligned}
x_{1,o} &= p_o q_o & x_{4,o} &= (1-p_o)(1-q_o) \\
x_{2,o} &= p_o(1-q_o) & D_o &= 0 \\
x_{3,o} &= q_o(1-p_o)
\end{aligned} \tag{3.10}$$

where p_o and q_o are the initial frequencies of the genes A and B, respectively. If the initial gene frequencies are the same, i.e. $q_o = p_o$, the relations given in (3.10) may be written in the form

$$\begin{aligned}
x_{1,o} &= p_o^2 & x_{4,o} &= (1-p_o)^2 \\
x_{2,o} &= x_{3,o} = p_o(1-p_o) & D_o &= 0
\end{aligned} \tag{3.11}$$

When selection is applied the population will no longer be in Hardy-Weinberg equilibrium. If we let $D_o = 0$ in (3.6) we get

$$\begin{aligned}
x_{1,n} &= \frac{1}{1-S+rS} \left[rS \left(\frac{1}{s^2} \right)^n + (1-S) \left(\frac{1-r}{S} \right)^n \right] x_{1,o} \\
x_{2,n} &= \frac{(1-S) \left[1-(1-r)^n \right] + rS(1-S^{-n})}{(1-S+rS) s^n} x_{1,o} + \frac{1}{s^n} x_{2,o} \\
x_{3,n} &= \frac{(1-S) \left[1-(1-r)^n \right] + rS(1-S^{-n})}{(1-S+rS) s^n} x_{1,o} + \frac{1}{s^n} x_{3,o} \\
D_n &= - \frac{1-S}{1-S+rS} \left[\left(\frac{1}{s^2} \right)^n - \left(\frac{1-r}{S} \right)^n \right] x_{1,o} .
\end{aligned} \tag{3.12}$$

Using the relations given by (3.10) we find

$$\begin{aligned}
x_{1,n} &= \frac{1}{1-S+rS} \left[rS \left(\frac{1}{s^2} \right)^n + (1-S) \left(\frac{1-r}{S} \right)^n \right] p_o q_o \neq p_n q_n \\
x_{2,n} &= \frac{(1-S) \left[1-(1-r)^n \right] + rS(1-S^{-n})}{(1-S+rS) s^n} p_o q_o + \frac{1}{s^n} p_o (1-q_o) \neq p_n (1-q_n) \\
x_{3,n} &= \frac{(1-S) \left[1-(1-r)^n \right] + rS(1-S^{-n})}{(1-S+rS) s^n} p_o q_o + \frac{1}{s^n} q_o (1-p_o) \neq q_n (1-p_n) \\
D_n &= \frac{1-S}{1-S+rS} \left[\left(\frac{1}{s^2} \right)^n - \left(\frac{1-r}{S} \right)^n \right] p_o q_o ,
\end{aligned}$$

where p_n and q_n are given by (3.9).

The change in the gametic frequencies will continue until $(1-S) \geq x_{4,n}$. Replacing $x_{4,n}$ by $(1-x_{1,n} - x_{2,n} - x_{3,n})$ we get $(1-S) \geq [1-(x_{1,n} + x_{2,n} + x_{3,n})]$. Replacing $x_{i,n}$ ($i = 1,2,3$) by their values from (3.6) we find

$$\begin{aligned}
(1-S) &\geq \left\{ 1 - \frac{[(rS)S^{-n} + (1-S)(1-r)^n]}{(1-S+rS) S^n} \right. \\
&\quad + \frac{2(1-S) [1-(1-r)^n] + 2rS(1-S^{-n})}{(1-S+rS) S^n} \left. \right\} x_{1,o} \\
&\quad - \frac{1}{S^n} (x_{2,o} + x_{3,o}) \\
&\quad - \frac{rS}{(1-S+rS) S^n} [S^{-n} - (1-r)^n] D_o \}^2 \\
&= \left\{ 1 - \frac{(1-S) [2-(1-r)^n] + rS(2-S^{-n})}{(1-S+rS) S^n} x_{1,o} \right. \\
&\quad \left. - \frac{1}{S^n} (x_{2,o} + x_{3,o}) - \frac{rS[S^{-n} - (1-r)^n]}{(1-S+rS) S^n} D_o \right\}^2 . \quad (3.13)
\end{aligned}$$

It is difficult to solve the above inequality for n . When $D_o = 0$, the inequality (3.13) reduces to

$$(1-S) \geq \left\{ 1 - \frac{(1-S) [2-(1-r)^n] + rS(2-S^{-n})}{(1-S+rS) S^n} x_{1,o} - \frac{1}{S^n} (x_{2,o} + x_{3,o}) \right\}^2 \quad (3.14)$$

For the case when $D_o = 0$ and $r = 0$, (3.13) reduces to

$$(1-S) \geq \left[1 - \frac{1}{S^n} (x_{1,o} + x_{2,o} + x_{3,o}) \right]^2 . \quad (3.15)$$

Using (3.15), the progress under the first stage continues until

$$n \geq \frac{1}{(-\ln S)} [\ln(1-\sqrt{1-S}) - \ln(x_{1,o} + x_{2,o} + x_{3,o})] . \quad (3.16)$$

Thus for a given S and initial frequencies $x_{1,o}$, $x_{2,o}$, $x_{3,o}$, the number of generations required for the first stage may be obtained from (3.13) or

approximately from (3.16) if $D_0 = 0$. Using the exact solutions (equation 3.13) with $x_{1,0} = .01$, $x_{2,0} = x_{3,0} = .09$ and $D_0 = 0$, the number of generations required for the first stage and the boundary values of x_4 and D , denoted by x_4^* and D^* up to which the first stage is operative, are given in Table 15 for different values of S and r .

Table 15. Number of generations required for the first stage and boundary values x_4^* and D^* .

	Selection S			
	.95	.9	.7	.5
Linkage $r = .50$	31	14	3	1
.30	31	14	3	1
.10	31	14	3	1
.05	29	13	3	1
x_4^*	.224	.340	.450	.620
D^*	0.0	-.028	-.052	-.020

Using equation (3.16) when $r = 0$ the number required for different values of S is

S	.95	.90	.70	.50
n	28	13	3	1 .

A comparison between the number of generations above and those given in Table 15 indicates that, in the present case, the effect of linkage is small.

$$2. \quad x_{4,n}^2 \leq (1-S) < x_{4,n}^2 + 2x_{2,n} x_{4,n} + 2x_{3,n} x_{4,n}.$$

In this stage all of the aabb and some of the Aabb and aaBb individuals are culled. Thus there are Aabb and aaBb among the individuals saved.

Two cases will be considered, one in which $x_2 = x_3$ and the second in which $x_2 \neq x_3$.

Case 1. $x_2 = x_3$:

Under this case the proportion culled will be equally divided among the genotypes Aabb and aaBb. The gametic frequencies in the next generation are given by

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} (x_{1,n} - rD_n) \\ x_{2,n+1} &= \frac{1}{S} \left[x_{2,n} - \frac{1}{4} (1-S - x_{4,n}^2) + rD_n \right] \\ x_{3,n+1} &= \frac{1}{S} \left[x_{3,n} - \frac{1}{4} (1-S - x_{4,n}^2) + rD_n \right] \\ x_{4,n+1} &= \frac{1}{S} \left[x_{4,n} - \frac{1}{2} (1-S + x_{4,n}^2) - rD_n \right]. \end{aligned} \tag{3.17}$$

If either D or r is small, we have approximately

$$\begin{aligned} x_{1,n} &= \frac{1}{S^n} x_{1,0} \\ x_{2,n} &= \frac{1}{S^n} x_{2,0} - \frac{1}{4} \sum_{i=0}^{n-1} S^{-n+i} (1-S-x_{4,i}^2) \\ x_{3,n} &= \frac{1}{S^n} x_{3,0} - \frac{1}{4} \sum_{i=0}^{n-1} S^{-n+i} (1-S-x_{4,i}^2) \\ x_{4,n} &= \frac{1}{S^n} x_{4,0} - \frac{1}{2} \sum_{i=0}^{n-1} S^{-n+i} (1-S+x_{4,i}^2). \end{aligned} \tag{3.18}$$

The gametic disequilibrium in the (n+1)th generation is given by

$$D_{n+1} = x_{1,n+1} x_{4,n+1} - x_{2,n+1} x_{3,n+1} .$$

Using equations (3.18) we find

$$D_{n+1} = \frac{1}{S^2} (1-rS) D_n - \frac{1}{S^2} f_n(x_{i,n}, S) , \quad (3.19)$$

where

$$\begin{aligned} f_n(x_{i,n}, S) = & \frac{1-S}{4} [2x_{1,n} - x_{2,n} - x_{3,n} + \frac{1}{4} (1-S)] \\ & + \frac{1}{4} x_{4,n}^2 [1 + x_{1,n} - x_{4,n} - \frac{1}{4} (2-2S-x_{4,n}^2)] . \end{aligned} \quad (3.20)$$

So, as long as the inequality in S holds,

$$D_n = \left(\frac{1-rS}{S^2}\right)^n D_0 - \frac{1}{S^2} \sum_{j=0}^{n-1} \left(\frac{1-rS}{S^2}\right)^{n-j-1} f_j(x_{i,j}, S) , \quad (3.21)$$

where $f_j(x_{i,j}, S)$ is given by (3.20). Equation (3.21) shows that selection may build up gametic disequilibrium during the second stage even when $D_0 = 0$. It is seen from the constant coefficient in (3.21) which involves S and r that gametic disequilibrium will be higher under strong selection and tight linkage.

The new gene frequencies are given by

$$p_{n+1} = x_{1,n+1} + x_{2,n+1}$$

$$q_{n+1} = x_{1,n+1} + x_{3,n+1}$$

where p and q are the frequencies of the A and B genes, respectively.

Using (3.17), we find

$$\begin{aligned}
 p_{n+1} &= \frac{1}{S} p_n - \frac{1}{4S} (1-S-x_{4,n}^2) \\
 q_{n+1} &= \frac{1}{S} q_n - \frac{1}{4S} (1-S-x_{4,n}^2) .
 \end{aligned}
 \tag{3.22}$$

So, as long as the inequality on S holds,

$$\begin{aligned}
 p_n &= \frac{1}{S^n} p_o - \frac{1}{4} \sum_{i=0}^{n-1} S^{-n+i} (1-S-x_{4,i}^2) \\
 q_n &= \frac{1}{S^n} q_o - \frac{1}{4} \sum_{i=0}^{n-1} S^{-n+i} (1-S-x_{4,i}^2) .
 \end{aligned}
 \tag{3.23}$$

Equations (3.23) indicate that during the second stage gene frequencies might slightly be affected by linkage through x_4 . Because during the second stage $(1-S) > x_{4,n}^2$, the rate of increase in gene frequencies will be reduced by the second term on the right-hand side of (3.23). Consider now the difference

$$\begin{aligned}
 q_n - p_n &= (x_{1,n} + x_{3,n}) - (x_{1,n} + x_{2,n}) \\
 &= x_{3,n} - x_{2,n} .
 \end{aligned}$$

Using (3.23), we find

$$\begin{aligned}
 q_n - p_n &= \frac{1}{S^n} (q_o - p_o) \\
 &= \frac{1}{S^n} (x_{3,o} - x_{2,o}) .
 \end{aligned}
 \tag{3.24}$$

Equation (3.24) shows that if the initial frequencies $x_{2,o} = x_{3,o}$ or $p_o = q_o$, gene frequencies will remain equal during the second stage. If $x_{2,o} \neq x_{3,o}$ the difference $x_{3,n} - x_{2,n}$ will be increased by selection.

In order to continuize the change in the gametic frequencies during this stage we may write approximately

$$\Delta x_{i,n} = x_{i,n+1} - x_{i,n} = \frac{dx_i}{dn}.$$

Using equations (3.17) and assuming that either D or r is small, we find

$$\begin{aligned} \ln x_{1,n} &= (\ln x_{1,0}) + n(1-S)/S \\ 2 \tan^{-1} \left(\frac{1-S-x_{4,n}}{\sqrt{S(1-S)}} \right) &= 2 \tan^{-1} \left(\frac{1-S-x_{4,0}}{\sqrt{S(1-S)}} \right) + n \sqrt{(1-S)/S}. \end{aligned} \quad (3.25)$$

Thus the function $(\ln x_1)$ is expected to change linearly with generations with slope given by $\hat{\beta}_2 = (1-S)/S$.

Case 2. $x_2 \neq x_3$:

In this case the proportion culled from the Aabb and aaBb genotypes will be proportional to their frequencies at the time of selection. The array among the culled individuals in generation n is

$$\frac{2x_{2,n}x_{4,n} \text{ Aabb} + 2x_{3,n}x_{4,n} \text{ aaBb}}{2x_{2,n}x_{4,n} + 2x_{3,n}x_{4,n}} = \frac{x_{2,n} \text{ Aabb} + x_{3,n} \text{ aaBb}}{x_{2,n} + x_{3,n}}.$$

The array among the saved individuals in the same generation is

$$\begin{aligned} &\frac{1}{S} \{ x_{1,n}^2 \text{ AABB} + 2x_{1,n}x_{3,n} \text{ AaBB} + 2x_{1,n}x_{2,n} \text{ AABb} \\ &+ 2x_{2,n}x_{3,n} \text{ Ab|aB} + 2x_{1,n}x_{4,n} \text{ AB|ab} + x_{3,n}^2 \text{ aaBB} + x_{2,n}^2 \text{ AAbb} \\ &+ [2x_{3,n}x_{4,n} - (1-S-x_{4,n}^2) \frac{x_{3,n}}{(x_{2,n} + x_{3,n})}] \text{ aaBb} \\ &+ [2x_{2,n}x_{4,n} - (1-S-x_{4,n}^2) \frac{x_{2,n}}{(x_{2,n} + x_{3,n})}] \text{ Aabb} \}. \end{aligned}$$

Thus we have

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} [x_{1,n}^2 + x_{1,n}x_{3,n} + x_{1,n}x_{2,n} + (1-r)x_{1,n}x_{4,n} + r x_{2,n}x_{3,n}] \\ &= \frac{1}{S} (x_{1,n} - r D_n) \end{aligned}$$

$$\begin{aligned} x_{2,n+1} &= \frac{1}{S} [x_{1,n}x_{2,n} + (1-r)x_{2,n}x_{3,n} + r x_{1,n}x_{4,n} \\ &\quad + x_{2,n}^2 + x_{2,n}x_{4,n} - \frac{1}{2} (1-S-x_{4,n}^2) \frac{x_{2,n}}{(x_{2,n} + x_{3,n})}] \\ &= \frac{1}{S} [x_{2,n} + r D_n - \frac{1}{2} (1-S-x_{4,n}^2) \frac{x_{2,n}}{(x_{2,n} + x_{3,n})}] \end{aligned}$$

(3.26)

$$\begin{aligned} x_{3,n+1} &= \frac{1}{S} [x_{1,n}x_{3,n} + (1-r)x_{2,n}x_{3,n} + r x_{1,n}x_{4,n} \\ &\quad + x_{3,n}^2 + x_{3,n}x_{4,n} - \frac{1}{2} (1-S-x_{4,n}^2) \frac{x_{3,n}}{(x_{2,n} + x_{3,n})}] \\ &= \frac{1}{S} [x_{3,n} + r D_n - \frac{1}{2} (1-S-x_{4,n}^2) \frac{x_{3,n}}{(x_{2,n} + x_{3,n})}] \end{aligned}$$

$$\begin{aligned} x_{4,n+1} &= \frac{1}{S} [r x_{2,n}x_{3,n} + (1-r)x_{1,n}x_{4,n} + x_{3,n}x_{4,n} \\ &\quad - \frac{1}{2} (1-S-x_{4,n}^2) \frac{x_{3,n}}{(x_{2,n} + x_{3,n})} + x_{2,n}x_{4,n} \\ &\quad - \frac{1}{2} (1-S-x_{4,n}^2) \frac{x_{2,n}}{(x_{2,n} + x_{3,n})}] \\ &= \frac{1}{S} [x_{4,n}(2-x_{4,n}) - r D_n - \frac{1}{2} (1-S)] \end{aligned}$$

If either D or r is small, we have approximately

$$x_{1,n} = \frac{1}{s^n} x_{1,0}$$

$$x_{2,n} = \frac{1}{s^n} x_{2,0} - \frac{1}{2} \sum_{i=0}^{n-1} s^{-n+i} (1-s-x_{4,i}^2) \frac{x_{2,i}}{(x_{2,i} + x_{3,i})} \quad (3.27)$$

$$x_{3,n} = \frac{1}{s^n} x_{3,0} - \frac{1}{2} \sum_{i=0}^{n-1} s^{-n+i} (1-s-x_{4,i}^2) \frac{x_{3,i}}{(x_{2,i} + x_{3,i})}$$

$$x_{4,n} = \frac{1}{s^n} x_{4,0} - \frac{1}{2} \sum_{i=0}^{n-1} s^{-n+i} (1-s+x_{4,i}^2) .$$

The gametic disequilibrium in the $(n+1)$ th generation is given by

$$D_{n+1} = x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1} .$$

Using (3.26), we find

$$D_{n+1} = \frac{1}{s^2} (1-rs) D_n + \frac{1}{s^2} x_{4,i}^2 D_n - \frac{1}{s^2} f_n(x_i, s) , \quad (3.28)$$

where

$$\begin{aligned} f_n(x_i, s) &= \frac{(1-s-x_{4,i}^2)}{(x_{2,n} + x_{3,n})^2} \left[\frac{1}{4} (1-s-x_{4,n}^2) - (x_{2,n} + x_{3,n}) \right] \\ &\quad + \frac{1}{2} x_{1,n} (1-s+x_{4,n}^2) . \end{aligned} \quad (3.29)$$

It is seen from (3.28) that selection may build up gametic disequilibrium even when the initial population is in gametic equilibrium, i.e. $D_0 = 0$.

The new frequencies of the A and B genes are given by

$$p_{n+1} = x_{1,n+1} + x_{2,n+1}$$

$$q_{n+1} = x_{1,n+1} + x_{3,n+1} ,$$

respectively. Using (3.26), we find

$$p_{n+1} = \frac{1}{S} p_n - \frac{1}{2S} (1-S-x_{4,n}^2) \frac{x_{2,n}}{(x_{2,n} + x_{3,n})} \quad (3.30)$$

$$q_{n+1} = \frac{1}{S} q_n - \frac{1}{2S} (1-S-x_{4,n}^2) \frac{x_{3,n}}{(x_{2,n} + x_{3,n})} .$$

So, as long as the inequality on S holds,

$$p_n = \frac{1}{S^n} p_0 - \frac{1}{2} \sum_{i=0}^{n-1} S^{-n+i} (1-S-x_{4,i}^2) \frac{x_{2,i}}{(x_{2,i} + x_{3,i})} \quad (3.31)$$

$$q_n = \frac{1}{S^n} q_0 - \frac{1}{2} \sum_{i=0}^{n-1} S^{-n+i} (1-S-x_{4,i}^2) \frac{x_{3,i}}{(x_{2,i} + x_{3,i})} .$$

As in Case 1 above, equations (3.31) indicate that gene frequencies may depend on linkage during the second stage through the gametic frequencies.

We now consider the difference

$$\begin{aligned} q_{n+1} - p_{n+1} &= (x_{1,n+1} + x_{3,n+1}) - (x_{1,n+1} + x_{2,n+1}) \\ &= x_{3,n+1} - x_{2,n+1} . \end{aligned}$$

Using (3.26), we find

$$q_{n+1} - p_{n+1} = \frac{1}{S} (x_{3,n} - x_{2,n}) \left[1 - \frac{(1-S-x_{4,n}^2)}{2(x_{2,n} + x_{3,n})} \right]$$

$$= \frac{1}{S^{n+1}} (q_o - p_o) \prod_{i=0}^n \left[1 - \frac{1-S-x_{4,i}^2}{2(x_{2,i} + x_{3,i})} \right] \quad (3.32)$$

It is seen from the inequality on S for the second stage that the product term given by the second term on the right-hand side of (3.32) is greater than zero. Equation (3.32) indicates that, when $q_o = p_o$ or $x_{2,o} = x_{3,o}$, the difference $q_n - p_n = 0$ in every generation. When $q_o \neq p_o$ or $x_{2,o} \neq x_{3,o}$, the difference $q_n - p_n$ will be increased by selection, and the rate of increase in the second stage is greater than that in the first stage.

It is seen from (3.17) and (3.26) that the change in the gametic frequencies will continue to the next stage.

$$3. \quad x_{4,n}^2 + 2x_{2,n}x_{4,n} + 2x_{3,n}x_{4,n} \leq (1-S) < 1 \\ - (x_{1,n}^2 + 2x_{1,n}x_{2,n} + 2x_{1,n}x_{3,n}).$$

The coupling and repulsion double heterozygotes are assumed to have the same fitness as $AAbb$ and $aaBB$ zygotes. In this stage all of the $aabb$, $Aabb$ and $aaBb$ and some of the $AAbb$, $aaBB$ and $AaBb$ individuals are culled.

In this case the proportion culled from the genotypes $AAbb$, $aaBB$, AB/ab and Ab/aB will be proportional to their frequencies at the time of selection. The array among the culled individuals in generation n is

$$\frac{x_{2,n}^2 AAbb + x_{3,n}^2 aaBB + 2x_{1,n}x_{4,n} AB/ab + 2x_{2,n}x_{3,n} Ab/aB}{x_{2,n}^2 + x_{3,n}^2 + 2x_{1,n}x_{4,n} + 2x_{2,n}x_{3,n}}$$

The array among the saved individuals in the same generation is

$$\begin{aligned}
& \frac{1}{S} \{ x_{1,n}^2 \text{AABB} + 2x_{1,n}x_{3,n} \text{AaBB} + 2x_{1,n}x_{2,n} \text{AABb} \\
& + [2x_{2,n}x_{3,n} - (1-S-x_{4,n}^2 - 2x_{2,n}x_{4,n} - 2x_{3,n}x_{4,n}) \frac{2x_{2,n}x_{3,n}}{A}] \\
& \text{Ab/aB} + [2x_{1,n}x_{4,n} - (1-S-x_{4,n}^2 - 2x_{2,n}x_{4,n} - 2x_{3,n}x_{4,n}) \\
& \frac{2x_{1,n}x_{4,n}}{A}] \text{AB/ab} + [x_{3,n}^2 - (1-S-x_{4,n}^2 - 2x_{2,n}x_{4,n} - 2x_{3,n}x_{4,n}) \frac{x_{3,n}^2}{A}] \text{aaBB} \\
& + [x_{2,n}^2 - (1-S-x_{4,n}^2 - 2x_{2,n}x_{4,n} - 2x_{3,n}x_{4,n}) \frac{x_{2,n}^2}{A}] \text{AAbb} \} ,
\end{aligned}$$

where

$$A = x_{2,n}^2 + x_{3,n}^2 + 2x_{1,n}x_{4,n} + 2x_{2,n}x_{3,n} . \quad (3.33)$$

Thus we have

$$\begin{aligned}
x_{1,n+1} &= \frac{1}{S} \{ x_{1,n}^2 + x_{1,n}x_{2,n} + x_{1,n}x_{3,n} + (1-r)x_{1,n}x_{4,n} \\
& - (1-r) \frac{x_{1,n}x_{4,n}}{A} (1-S-x_{4,n}^2 - 2x_{2,n}x_{4,n} - 2x_{3,n}x_{4,n}) \\
& + r x_{2,n}x_{3,n} - 4 \frac{x_{2,n}x_{3,n}}{A} (1-S-x_{4,n}^2 - 2x_{2,n}x_{4,n} - 2x_{3,n}x_{4,n}) \} \\
&= \frac{1}{S} \{ x_{1,n} - \frac{1}{A} r D_n [S-x_{1,n}(x_{1,n}+2x_{2,n}+2x_{3,n})] \\
& - \frac{1}{A} x_{1,n}x_{4,n} (1-S-x_{4,n}^2 - 2x_{2,n}x_{4,n} - 2x_{3,n}x_{4,n}) \} \quad (3.34)
\end{aligned}$$

$$\begin{aligned}
x_{2,n+1} &= \frac{1}{S} \{x_{1,n}x_{2,n} + (1-r)x_{2,n}x_{3,n} - \frac{1}{A} (1-r)x_{2,n}x_{3,n} \\
&\quad (1-S-x_{4,n}^2-2x_{2,n}x_{4,n}-2x_{3,n}x_{4,n}) + r x_{1,n}x_{4,n} \\
&\quad - \frac{1}{A} r x_{1,n}x_{4,n} (1-S-x_{4,n}^2-2x_{2,n}x_{4,n}-2x_{3,n}x_{4,n}) \\
&\quad + x_{2,n}^2 - \frac{1}{A} x_{2,n}^2 (1-S-x_{4,n}^2-2x_{2,n}x_{4,n}-2x_{3,n}x_{4,n})\} \\
&= \frac{1}{S} \{x_{2,n}(1-x_{4,n}) + \frac{1}{A} rD_n [S - x_{1,n}(x_{1,n}+2x_{2,n}+2x_{3,n})] \\
&\quad - \frac{1}{A} x_{2,n} [x_{3,n} + x_{2,n}(1-S-x_{4,n}^2-2x_{2,n}x_{4,n}-2x_{3,n}x_{4,n})]\}
\end{aligned} \tag{3.35}$$

$$\begin{aligned}
x_{3,n+1} &= \frac{1}{S} \{x_{1,n}x_{3,n} + (1-r)x_{2,n}x_{3,n} - \frac{1}{A} (1-r)x_{2,n}x_{3,n} \\
&\quad (1-S-x_{4,n}^2-2x_{2,n}x_{4,n}-2x_{3,n}x_{4,n}) + r x_{1,n}x_{4,n} \\
&\quad - \frac{1}{A} r x_{1,n}x_{4,n} (1-S-x_{4,n}^2-2x_{2,n}x_{4,n}-2x_{3,n}x_{4,n}) \\
&\quad + x_{3,n}^2 - \frac{1}{A} x_{3,n}^2 (1-S-x_{4,n}^2-2x_{2,n}x_{4,n}-2x_{3,n}x_{4,n})\} \\
&= \frac{1}{S} \{x_{3,n}(1-x_{4,n}) + \frac{1}{A} rD_n [S-x_{1,n}(x_{1,n}+2x_{2,n}+2x_{3,n})] \\
&\quad - \frac{1}{A} x_{3,n} [x_{2,n} + x_{3,n}(1-S-x_{4,n}^2-2x_{2,n}x_{4,n}-2x_{3,n}x_{4,n})]\}
\end{aligned}$$

$$\begin{aligned}
x_{4,n+1} &= \frac{1}{S} \{r x_{2,n}x_{3,n} - \frac{1}{A} r x_{2,n}x_{3,n} (1-S-x_{4,n}^2-2x_{2,n}x_{4,n}-2x_{3,n}x_{4,n}) \\
&\quad + (1-r)x_{1,n}x_{4,n} - \frac{1}{A} (1-r) x_{1,n}x_{4,n} (1-S-x_{4,n}^2 \\
&\quad - 2x_{2,n}x_{4,n}-2x_{3,n}x_{4,n})\} \\
&= \frac{1}{S} A^{-1}(x_{1,n}x_{4,n}-rD_n) [S - x_{1,n}(x_{1,n}+2x_{2,n}+2x_{3,n})] ,
\end{aligned}$$

where A is given by (3.34).

The new frequencies of the A and B genes are given by

$$p_{n+1} = x_{1,n+1} + x_{2,n+1}$$

$$q_{n+1} = x_{1,n+1} + x_{3,n+1}$$

respectively. Using (3.35), we find

$$\begin{aligned} p_{n+1} &= \frac{1}{S} p_n - \frac{1}{S} \left[\frac{V}{A} (x_{1,n} x_{4,n} + x_{2,n}^2) + x_{2,n} \left(\frac{1}{A} x_{3,n} + x_{4,n} \right) \right] \\ q_{n+1} &= \frac{1}{S} q_n - \frac{1}{S} \left[\frac{V}{A} (x_{1,n} x_{4,n} + x_{3,n}^2) + x_{3,n} \left(\frac{1}{A} x_{2,n} + x_{4,n} \right) \right], \end{aligned} \quad (3.36)$$

where

$$V = (1 - S - x_{4,n}^2 - 2x_{2,n} x_{4,n} - 2x_{3,n} x_{4,n}) \quad (3.37)$$

and A is defined in (3.34). So, as long as the inequality on S holds

$$\begin{aligned} p_n &= \frac{1}{S^n} p_0 - \sum_{i=0}^{n-1} S^{-n+i} f_i \\ q_n &= \frac{1}{S^n} q_0 - \sum_{i=0}^{n-1} S^{-n+i} g_i, \end{aligned} \quad (3.38)$$

where

$$f_i = \frac{V}{A} (x_{1,i} x_{4,i} + x_{2,i}^2) + x_{2,i} \left(\frac{1}{A} x_{3,i} + x_{4,i} \right) \quad (3.39)$$

$$g_i = \frac{V}{A} (x_{1,i} x_{4,i} + x_{3,i}^2) + x_{3,i} \left(\frac{1}{A} x_{2,i} + x_{4,i} \right) \quad (3.40)$$

with A and V defined as in (3.34) and (3.35), respectively. Equation (3.38) show that gene frequencies may be affected by linkage during the third stage. It is seen from (3.31) and (3.38) that the rate of change in p_n will be slower in the third stage than in the second stage if

$$\frac{1}{2} (1 - S - x_{4,i}^2) \frac{x_{2,i}}{(x_{2,i} + x_{3,i})} < f_i ,$$

where f_i is defined in (3.39). We now consider the difference

$$\begin{aligned} q_{n+1} - p_{n+1} &= (x_{1,n+1} + x_{3,n+1}) - (x_{1,n+1} + x_{2,n+1}) \\ &= x_{3,n+1} - x_{2,n+1} . \end{aligned}$$

Using (3.35), we find

$$\begin{aligned} q_{n+1} - p_{n+1} &= \frac{1}{S}(x_{3,n} - x_{2,n}) \left[1 - x_{4,n} - \frac{V}{A}(x_{2,n} + x_{3,n}) \right] \\ &= \frac{1}{S}(q_n - p_n) \left[1 - x_{4,n} - \frac{V}{A}(x_{2,n} + x_{3,n}) \right] , \end{aligned} \quad (3.41)$$

where A and V are defined in (3.34) and (3.37), respectively. So, as long as the inequality on S holds

$$\begin{aligned} q_n - p_n &= \frac{1}{S^n} (x_{3,o} - x_{2,o}) \prod_{i=0}^{n-1} \left[1 - x_{4,i} - \frac{V}{A}(x_{2,i} + x_{3,i}) \right] \\ &= \frac{1}{S^n} (q_o - p_o) \prod_{i=0}^{n-1} \left[1 - x_{4,i} - \frac{V}{A}(x_{2,i} + x_{3,i}) \right] \end{aligned} \quad (3.42)$$

where A and V are given by (3.34) and (3.37), respectively. It is seen from (3.42) that if the initial frequencies $q_o = p_o$ or $x_{2,o} = x_{3,o}$, the gene frequencies will remain equal in every generation. When $q_o \neq p_o$ or $x_{2,o} \neq x_{3,o}$, the difference $q_n - p_n$ will be increased by selection. When $x_{2,o} \neq x_{3,o}$, it is seen from (3.32) and (3.42) that the rate of increase in the difference between the frequencies of the two genes will be higher in the third stage than in the second stage if

$$\frac{1-S-x_{4,i}^2}{2(x_{2,i}+x_{3,i})} > x_{4,i} + \frac{V}{A} (x_{2,i} + x_{3,i}) ,$$

where A and V are defined in (3.34) and (3.37), respectively.

$$4. \quad 1 - (x_{1,n}^2 + 2x_{1,n}x_{2,n} + 2x_{1,n}x_{3,n}) \leq (1-S) < 1 - x_{1,n}^2.$$

In this stage we have $x_1 + x_2 + x_3 = 1$ and $x_4 = 0$. It is seen in this stage that all the individuals AABB and some of the AABb and AaBB will be saved. Two cases will be considered, one in which $x_2 = x_3$ and the second in which $x_2 \neq x_3$.

Case 1. $x_2 = x_3$:

In this case the proportion culled will be equally divided among the genotypes AABb and AaBB. The gametic frequencies in the next generation are given by

$$\begin{aligned} x_{1,n+1} &= \frac{1}{2S} (S + x_{1,n}^2) \\ x_{2,n+1} &= \frac{1}{4S} (S - x_{1,n}^2) \\ x_{3,n+1} &= \frac{1}{4S} (S - x_{1,n}^2) \\ x_{4,n+1} &= 0 . \end{aligned} \tag{3.43}$$

It is clear from the above equations that the gametic frequencies change independently of linkage during this stage. It is seen also that gene frequency is independent of linkage for we have

$$p_{n+1} = \frac{3}{4} + \frac{1}{4} S^{-1} x_{1,n}^2 \tag{3.44}$$

and will approach unity as x_1 approaches \sqrt{S} . So that the final equilibrium for gametic or gene frequencies for the additive model occurs with $D = 0$.

In order to continuize the change in the gametic frequencies during this stage we write $\Delta x_{1,n} = x_{1,n+1} - x_{1,n} \doteq (dx_1/dn)$. Using equation (3.43) we find

$$2 \tan^{-1} \frac{x_{1,n} - S}{\sqrt{S(1-S)}} = 2 \tan^{-1} \frac{x_{1,0} - S}{\sqrt{S(1-S)}} + n \sqrt{(1-S)/S}, \quad (3.45)$$

indicating a linear relationship between the function of the gametic frequency given by the left-hand side of (3.45) with generations, with slope given by $\hat{\beta}_4 = \sqrt{(1-S)/S}$. Figure 4 shows the change of this function with n for different values of S and r using the exact numerical solutions. The estimates of the slope ($\hat{\beta}_4$) are in agreement with the exact slope (β), Table 28, Appendix B.

Case 2. $x_2 \neq x_3$:

In this case the proportion culled from the AABb and AaBB genotypes will be proportional to their frequencies at the time of selection. The array among the culled individuals in generation n is

$$\frac{2x_{1,n}x_{2,n}^{AaBB} + 2x_{1,n}x_{3,n}^{AaBB}}{2x_{1,n}x_{2,n} + 2x_{1,n}x_{3,n}} = \frac{x_{2,n}^{AaBB} + x_{3,n}^{AaBB}}{(x_{2,n} + x_{3,n})}.$$

The array among the saved individuals in the same generation is

$$\begin{aligned} & \frac{1}{S} \{x_{1,n}^2 \text{AABB} + [2x_{1,n}x_{2,n} - (x_{1,n} + x_{1,n}x_{2,n} + x_{1,n}x_{2,n} - S) \\ & \quad \frac{x_{2,n}}{(x_{2,n} + x_{3,n})}] \text{AABb} + [2x_{1,n}x_{3,n} - (x_{1,n} + x_{1,n}x_{2,n} + x_{1,n}x_{3,n} - S) \\ & \quad \frac{x_{3,n}}{(x_{2,n} + x_{3,n})}] \text{AaBB}\} . \end{aligned}$$

Thus we have

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} \{x_{1,n}^2 + x_{1,n}x_{2,n} + x_{1,n}x_{3,n} - \frac{1}{2} (x_{1,n} + x_{1,n}x_{2,n} + x_{1,n}x_{3,n} - S)\} \\ &= \frac{1}{2S} (x_{1,n}^2 + S) \end{aligned}$$

$$\begin{aligned} x_{2,n+1} &= \frac{1}{S} \{x_{1,n}x_{2,n} - \frac{1}{2}(x_{1,n} + x_{1,n}x_{2,n} + x_{1,n}x_{3,n} - S) \frac{x_{2,n}}{(x_{2,n} + x_{3,n})}\} \\ &= \frac{1}{2S} \frac{x_{2,n}}{(x_{2,n} + x_{3,n})} (S - x_{1,n}^2) \end{aligned} \quad (3.46)$$

$$\begin{aligned} x_{3,n+1} &= \frac{1}{S} \{x_{1,n}x_{3,n} - \frac{1}{2}(x_{1,n} + x_{1,n}x_{2,n} + x_{1,n}x_{3,n} - S) \frac{x_{3,n}}{(x_{2,n} + x_{3,n})}\} \\ &= \frac{1}{2S} \frac{x_{3,n}}{(x_{2,n} + x_{3,n})} (S - x_{1,n}^2) \end{aligned}$$

$$x_{4,n+1} = 0$$

It is seen from (3.46) that the gametic frequencies change independently of linkage during the fourth stage. It is also seen that as x_1 approaches \sqrt{S} , the frequencies x_2 and x_3 will approach zero and we have at equilibrium $x_1 = 1$, $x_2 = x_3 = x_4 = 0$.

The gametic disequilibrium in the $(n+1)$ th generation is given by

$$D_{n+1} = x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1} .$$

Using (3.46), we find

$$D_{n+1} = -\frac{1}{4S} x_{2,n}x_{3,n} \left(\frac{S - x_{1,n}^2}{x_{2,n} + x_{3,n}} \right)^2 . \quad (3.47)$$

The gametic disequilibrium will therefore approach zero as x_1 approaches \sqrt{S} or either x_2 or x_3 approaches zero.

The new frequencies of the A and B genes are given by

$$p_{n+1} = x_{1,n+1} + x_{2,n+1}$$

$$q_{n+1} = x_{1,n+1} + x_{3,n+1} ,$$

respectively. We now consider the difference

$$q_{n+1} - p_{n+1} = x_{3,n+1} - x_{2,n+1} .$$

Using (3.46), we find

$$\begin{aligned} q_{n+1} - p_{n+1} &= \frac{1}{2S} (x_{3,n} - x_{2,n}) \left(\frac{S - x_{1,n}^2}{x_{2,n} + x_{3,n}} \right) \\ &= \frac{1}{2S} (q_n - p_n) \left(\frac{S - x_{1,n}^2}{x_{2,n} + x_{3,n}} \right) \end{aligned} \quad (3.48)$$

So, as long as the inequality on S holds

$$q_n - p_n = \frac{1}{(2S)^n} (q_0 - p_0) \prod_{i=0}^{n-1} \left(\frac{S - x_{1,i}^2}{x_{2,i} + x_{3,i}} \right) .$$

It is seen from (3.48) that at equilibrium or as x_1 approaches \sqrt{S} the frequencies of the two genes A and B will be equal.

$$5. \quad (1-S) \geq 1 - x_{1,n}^2 \text{ or } S \leq x_{1,n}^2 .$$

Since, in this stage, only AABB individuals are allowed to reproduce, we find

$$x_{1,n+1} = 1$$

$$x_{2,n+1} = x_{3,n+1} = 0$$

$$x_{4,n+1} = 0$$

It is clear from the above treatment that linkage might have some effect on the change in the gametic frequencies especially during the first three stages during which selection is building up negative gametic disequilibrium. However, the effect of linkage on the rate of change in gene frequencies is very small (equations 3.22 and 3.30). It is seen that the magnitude of the effect of linkage would depend on the magnitude of D , the gametic disequilibrium. Because selection operating on an additive gene may build up a very small amount of gametic disequilibrium (Figure 3), we would expect the magnitude of the effect of linkage to be small, the effect is relatively important for intermediate values of S . However, when selection is strong, its effect will overcome the effect of segregation. This has been pointed out by Fisher (1930) when examining the effect of natural selection using constant selective values for the genotypes.

We consider now the change in the genotypic mean of the population under the present additive model. Consider the mean phenotypic values given in Table 16.

Table 16. Mean phenotypic values of the genotypes.

	AA	Aa	aa
BB	$a_2 + b_2$	$a_1 + b_2$	$a_0 + b_2$
Bb	$a_2 + b_1$	$a_1 + b_1$	$a_0 + b_1$
bb	$a_2 + b_0$	$a_1 + b_0$	$a_0 + b_0$

If there are equal effects at the two loci we have $a_i = b_i$ ($i = 0, 1, 2$).
The population mean in generation n is given by

$$\begin{aligned}
W_n &= (a_2 + b_2)x_{1,n}^2 + 2(a_2 + b_1)x_{1,n}x_{2,n} + (a_2 + b_0)x_{2,n}^2 \\
&\quad + 2(a_1 + b_2)x_{1,n}x_{3,n} + 2(a_1 + b_1)(x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) \\
&\quad + 2(a_1 + b_0)x_{2,n}x_{4,n} + (a_0 + b_2)x_{3,n}^2 \\
&\quad + 2(a_0 + b_1)x_{3,n}x_{4,n} + (a_0 + b_0)x_{4,n}^2 \\
&= a_2(x_{1,n} + x_{2,n})^2 + 2a_1(x_{1,n} + x_{2,n})(x_{3,n} + x_{4,n}) \\
&\quad + a_0(x_{3,n} + x_{4,n})^2 + b_2(x_{1,n} + x_{3,n})^2 \\
&\quad + 2b_1(x_{1,n} + x_{3,n})(x_{2,n} + x_{4,n}) + b_0(x_{2,n} + x_{4,n})^2 \\
&= a_2p_n^2 + 2a_1p_n(1-p_n) + a_0(1-p_n)^2 + b_2q_n^2 \\
&\quad + 2b_1q_n(1-q_n) + b_0(1-q_n)^2,
\end{aligned}$$

where p_n and q_n are the frequencies of the genes A and B in generation n , respectively. It is seen from (3.9) that gene frequencies do not depend on linkage during the first stage. Consequently, the mean of the

population is expected to change independently of linkage. As the progress of the population proceeds under selection and during the second and the third stages it is seen from (3.22), (3.30) and (3.36) that gene frequencies may depend on linkage through the gametic frequencies. Accordingly, linkage may have a slight effect on the change in the population mean during the second and third stages.

Complete dominance at both loci. The mean phenotypic values for the genotypes are presented in Table 27, Appendix B. The numerical values in the body of the table determine only the order of the genotypes. The mathematics that follows does not depend on the particular values used. For this model the order of the genotypes is

$$aabb < aaBb = aaBB = Aabb = AAbb < AB/ab = Ab/aB = AABb = AaBB = AABB.$$

There are three possible stages.

$$1. \quad (1-S) < x_{4,n}^2.$$

This stage is the same as the first stage discussed above under the model with no dominance. Accordingly, equations (3.1) to (3.16) apply to this stage as well as the discussion on these equations. Figure 4 shows the exact linear change in the function of the gametic frequency in x_1 with generations. Table 29, Appendix B, gives the exact values for the slope β and their estimates $\hat{\beta}$ for different values of S and r .

$$2. \quad x_{4,n}^2 \leq (1-S) < x_{4,n}^2 + 2x_{2,n}x_{4,n} + 2x_{3,n}x_{4,n} + x_{2,n}^2 + x_{3,n}^2.$$

In this stage all of the aabb and some of the aaBb, aaBB, Aabb and AAbb individuals are culled. The array among the culled individuals that are not of genotype aabb in generation n is

$$\frac{1}{W_n} \{2x_{2,n}x_{4,n} \text{ Aabb} + 2x_{3,n}x_{4,n} \text{ aaBb} + x_{2,n}^2 \text{ AAbb} + x_{3,n}^2 \text{ aaBB}\} ,$$

where

$$W_n = 2x_{2,n}x_{4,n} + 2x_{3,n}x_{4,n} + x_{2,n}^2 + x_{3,n}^2 . \quad (3.49)$$

The array among the saved individuals in the same generation is

$$\begin{aligned} & \frac{1}{S} \{x_{1,n}^2 \text{ AABB} + 2x_{1,n}x_{3,n} \text{ AaBB} + 2x_{1,n}x_{2,n} \text{ AABb} \\ & + 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) \text{ AaBb} \\ & + [2x_{2,n}x_{4,n} - (1-S-x_{4,n}^2) 2x_{2,n}x_{4,n} W_n^{-1}] \text{ Aabb} \\ & + [2x_{3,n}x_{4,n} - (1-S-x_{4,n}^2) 2x_{3,n}x_{4,n} W_n^{-1}] \text{ aaBb} \\ & + [x_{2,n}^2 - (1-S-x_{4,n}^2) x_{2,n}^2 W_n^{-1}] \text{ AAbb} \\ & + [x_{3,n}^2 - (1-S-x_{4,n}^2) x_{3,n}^2 W_n^{-1}] \text{ aaBB}\} , \end{aligned}$$

where W_n is defined in (3.49).

Thus we find

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} [x_{1,n}^2 + x_{1,n}x_{3,n} + x_{1,n}x_{2,n} + (1-r) x_{1,n}x_{4,n} + r x_{2,n}x_{3,n}] \\ &= \frac{1}{S} (x_{1,n} - rD_n) \end{aligned}$$

$$\begin{aligned}
x_{2,n+1} &= \frac{1}{S} \left[x_{1,n} x_{2,n} + r x_{1,n} x_{4,n} + (1-r) x_{2,n} x_{3,n} + x_{2,n} x_{4,n} \right. \\
&\quad \left. - (1-S-x_{4,n}^2) x_{2,n} x_{4,n} W_n^{-1} + x_{2,n}^2 - (1-S-x_{4,n}^2) x_{2,n}^2 W_n^{-1} \right] \\
&= \frac{1}{S} \left[x_{2,n} + r D_n - x_{2,n} (x_{2,n} + x_{4,n}) (1-S-x_{4,n}^2) W_n^{-1} \right] \\
x_{3,n+1} &= \frac{1}{S} \left[x_{1,n} x_{3,n} + r x_{1,n} x_{4,n} + (1-r) x_{2,n} x_{3,n} + x_{3,n} x_{4,n} \right. \\
&\quad \left. - x_{3,n} x_{4,n} (1-S-x_{4,n}^2) W_n^{-1} + x_{3,n}^2 - (1-S-x_{4,n}^2) x_{3,n}^2 W_n^{-1} \right] \\
&= \frac{1}{S} \left[x_{3,n} + r D_n - x_{3,n} (x_{3,n} + x_{4,n}) (1-S-x_{4,n}^2) W_n^{-1} \right] \quad (3.50)
\end{aligned}$$

$$\begin{aligned}
x_{4,n+1} &= \frac{1}{S} \left[(1-r) x_{1,n} x_{4,n} + r x_{2,n} x_{3,n} + x_{2,n} x_{4,n} - x_{2,n} x_{4,n} \right. \\
&\quad \left. (1-S-x_{4,n}^2) W_n^{-1} + x_{3,n} x_{4,n} - x_{3,n} x_{4,n} (1-S-x_{4,n}^2) W_n^{-1} \right] \\
&= \frac{1}{S} \left[x_{4,n} (1-x_{4,n}) - r D_n - x_{4,n} (x_{2,n} + x_{3,n}) (1-S-x_{4,n}^2) W_n^{-1} \right]
\end{aligned}$$

where W_n is defined in (3.49) The new frequencies of the A and B genes are given by

$$p_{n+1} = x_{1,n+1} + x_{2,n+1}$$

$$q_{n+1} = x_{1,n+1} + x_{3,n+1},$$

respectively. Using (3.50), we find

$$\begin{aligned}
p_{n+1} &= \frac{1}{S} p_n - \frac{1}{S} x_{2,n} (x_{2,n} + x_{4,n}) (1-S-x_{4,n}^2) W_n^{-1} \\
q_{n+1} &= \frac{1}{S} q_n - \frac{1}{S} x_{3,n} (x_{3,n} + x_{4,n}) (1-S-x_{4,n}^2) W_n^{-1}, \quad (3.51)
\end{aligned}$$

where W_n is defined in (3.49). Equations (3.51) show that changes in the gene frequencies may depend on linkage through the gametic frequencies during the second stage. We now consider the difference

$$q_{n+1} - p_{n+1} = x_{3,n+1} - x_{2,n+1} .$$

Using (3.51), we find

$$q_{n+1} - p_{n+1} = \frac{1}{S} (q_n - p_n) [1 - (1-x_{1,n})(1-S-x_{4,n}^2) W_n^{-1}] \quad (3.52)$$

where W is defined in (3.49). So, as long as the inequality on S holds

$$q_n - p_n = \frac{1}{S^n} (q_0 - p_0) \prod_{i=0}^{n-1} [1 - (1-x_{1,i})(1-S-x_{4,i}^2) W_i^{-1}] . \quad (3.53)$$

Equation (3.53) shows that when $q_0 = p_0$ or $x_{2,0} = x_{3,0}$ the frequencies of the two genes A and B will remain equal in every generation. When

$x_{2,0} \neq x_{3,0}$, the difference $(x_{3,n} - x_{2,n})$ will be increased by selection.

From the defining inequality for this stage it is seen that the term

$[1 - (1-x_{1,i})(1-S-x_{4,i}^2) W_i^{-1}]$ appearing in (3.53) is greater than one.

Thus, when $x_{2,0} \neq x_{3,0}$, the rate of increase in the difference $(x_{3,n} - x_{2,n})$ by selection is higher in the second than in the first stage.

The gametic disequilibrium in the $(n+1)$ th generation is given by

$$\begin{aligned} D_{n+1} = & \frac{1}{S^2} \{1 - r + r [x_{4,n}^2 + (1-S-x_{4,n}^2) W_n^{-1} (x_{2,n}^2 + x_{3,n}^2 \\ & - (x_{2,n} + x_{3,n})(1-2x_{4,n}))]\} D_n - \frac{1}{S^2} \{x_{1,n}x_{4,n}^2 \\ & - (1-S-x_{4,n}^2) W_n^{-1} [2x_{2,n}x_{3,n}x_{4,n} + (1-S-x_{4,n}^2)x_{2,n}x_{3,n} \\ & (x_{2,n} + x_{4,n})(x_{3,n} + x_{4,n}) W_n^{-1}]\} , \end{aligned} \quad (3.54)$$

where W_n is defined in (3.49). Equation (3.54) indicates that selection may build up gametic disequilibrium during this stage.

The change in the gametic frequencies is given by

$$\Delta x_{i,n} = x_{i,n+1} - x_{i,n} \quad i = 1, 2, 3, 4 .$$

Using (3.50), we find

$$\begin{aligned} \Delta x_{1,n} &= \frac{1}{S} [(1-S) x_{1,n} - rD_n] \\ \Delta x_{2,n} &= \frac{1}{S} [(1-S) x_{2,n} + rD_n - x_{2,n}(x_{2,n} + x_{4,n})(1-S-x_{4,n}^2) W_n^{-1}] \\ \Delta x_{3,n} &= \frac{1}{S} [(1-S) x_{3,n} + rD_n - x_{3,n}(x_{3,n} + x_{4,n})(1-S-x_{4,n}^2) W_n^{-1}] \\ \Delta x_{4,n} &= \frac{1}{S} [(1-S-x_{4,n}^2) x_{4,n} - rD_n - x_{4,n}(x_{2,n} + x_{3,n})(1-S-x_{4,n}^2) W_n^{-1}] , \end{aligned} \quad (3.55)$$

where W_n is defined in (3.49). Equations (3.50) and (3.55) indicate that the gametic frequencies will continue to change up to the next stage.

$$3. \quad (1-S) \geq x_{4,n}^2 + 2x_{2,n}x_{4,n} + 2x_{3,n}x_{4,n} + x_{2,n}^2 + x_{3,n}^2 .$$

We find

$$\begin{aligned} x_{1,n+1} &= W_n^{-1} (x_{1,n} - rD_n) \\ x_{2,n+1} &= W_n^{-1} [x_{2,n}(x_{1,n} + x_{2,n}) + rD_n] \\ x_{3,n+1} &= W_n^{-1} [x_{3,n}(x_{1,n} + x_{3,n}) + rD_n] \\ x_{4,n+1} &= W_n^{-1} (x_{1,n}x_{4,n} - rD_n) , \end{aligned} \quad (3.56)$$

where

$$W_n = x_{1,n}(2-x_{1,n}) + 2x_{2,n}x_{3,n} . \quad (3.57)$$

Equations (3.56) indicate that linkage may have some effect on the changes in gametic frequencies. The change in gene frequency may also depend on linkage through the gametic frequencies, for we have

$$p_{n+1} = x_{1,n+1} + x_{2,n+1}$$

$$q_{n+1} = x_{1,n+1} + x_{3,n+1}$$

where p and q are the frequencies of the two genes A and B, respectively.

Using equations (3.56), we find

$$\begin{aligned} p_{n+1} &= W_n^{-1} (p_n x_{2,n} + x_{1,n}) \\ q_{n+1} &= W_n^{-1} (q_n x_{3,n} + x_{1,n}) , \end{aligned} \quad (3.58)$$

where

$$W_n = x_{1,n}(2-x_{1,n}) + 2x_{2,n}x_{3,n} .$$

We now consider the difference

$$q_{n+1} - p_{n+1} = x_{3,n+1} - x_{2,n+1} .$$

Using (3.56), we find

$$\begin{aligned} q_{n+1} - p_{n+1} &= (x_{3,n} - x_{2,n}) W_n^{-1} (1-x_{4,n}) \\ &= (q_n - p_n) W_n^{-1} (1-x_{4,n}) . \end{aligned} \quad (3.59)$$

So, as long as the inequality on S holds,

$$q_n - p_n = (q_0 - p_0) \prod_{i=0}^{n-1} W_i^{-1} (1-x_{4,i}) , \quad (3.60)$$

where

$$W_i = x_{1,i}(2-x_{1,i}) + 2x_{2,i}x_{3,i} .$$

Equation (3.60) shows that when $q_o = p_o$ or $x_{2,o} = x_{3,o}$ the frequencies of the two genes A and B will remain equal in every generation. The gametic disequilibrium in the $(n+1)$ th generation is given by

$$D_{n+1} = x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1} .$$

Using equations (3.56) we find

$$D_{n+1} = W_n^{-1} \left[\left(\frac{1}{2} - r \right) + \frac{1}{2} W_n^{-1} x_{1,n}^2 \right] D_n . \quad (3.61)$$

Equation (3.61) for the change in gametic disequilibrium indicates that, when $D_o = 0$, gametic equilibrium will be maintained in the population. When $D_o \neq 0$, loose linkage will reduce the amount of D. Thus the value of D will be increased when selection is strong and linkage is tight. The equation also indicates that selection may build up gametic disequilibrium even when linkage is loose, $r = \frac{1}{2}$.

In order to continuize the change in the gametic frequencies during this stage we write

$$\Delta x_{1,n} = x_{1,n+1} - x_{1,n} \doteq (dx_1/dn) .$$

Using equations (3.56) we find, approximately,

$$\frac{1}{1-x_{1,n}} - \ln(1-x_{1,n}) = \frac{1}{1-x_{1,o}} - \ln(1-x_{1,o}) + n, \quad (3.62)$$

indicating that the function of the gametic frequency given on the left-hand side of (3.62) is expected to change linearly with generations with

slope $\hat{\beta} = 1$, independent of selection. We note here that equation (3.62)

is similar to that obtained under the one locus case (equation 1.36).

Figure 4 shows the exact change of the function of x_1 with generations under different values of linkage and selection intensities when $p_0 = q_0 = .1$ and $D_0 = 0$. Linkage and degree of selection do not have profound effects on the change in the gametic frequencies. Of course, the degree of selection must lie in the specified range.

Overdominance for both loci with interaction. The mean phenotypic values of the genotypes are presented in Table 27, Appendix B. The numerical values in the body of the table determine only the order of the genotypes. The mathematics that follows does not depend on the particular values used. The order of the genotypes is

$$aabb = Aabb = AAbb = aaBb = aaBB = AABb = AaBB = AABB < AB/ab = Ab/aB.$$

This model involves interaction between the two loci. There are two possible stages of progress.

$$1. \quad (1-S) < 1 - 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n}).$$

In this stage all of the AaBb and some of the aabb, Aabb, AAbb, aaBb, aaBB, AABb, AaBB and AABB individuals are saved. The array among the culled individuals in generation n is

$$\frac{1}{1 - 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n})} \{ x_{4,n}^2 aabb + 2x_{3,n}x_{4,n} aABb + 2x_{2,n}x_{4,n} Aabb + x_{2,n}^2 AAbb + x_{3,n}^2 aaBB + 2x_{1,n}x_{2,n} AABb + 2x_{1,n}x_{3,n} AaBB + x_{1,n}^2 AABB \}.$$

The array among the saved individuals in the same generation is

$$\begin{aligned}
& \frac{1}{S} \{ x_{4,n}^2 (1 - \frac{1-S}{W_n}) aabb + 2x_{2,n}x_{4,n} (1 - \frac{1-S}{W_n}) Aabb \\
& + 2x_{3,n}x_{4,n} (1 - \frac{1-S}{W_n}) aaBb + x_{2,n}^2 (1 - \frac{1-S}{W_n}) AAab \\
& + x_{3,n}^2 (1 - \frac{1-S}{W_n}) aaBB + 2x_{1,n}x_{2,n} (1 - \frac{1-S}{W_n}) AABb \\
& + 2x_{1,n}x_{3,n} (1 - \frac{1-S}{W_n}) AaBB + x_{1,n}^2 (1 - \frac{1-S}{W_n}) AABB \\
& + 2x_{1,n}x_{4,n} AB/ab + 2x_{1,n}x_{3,n} Ab/aB \} ,
\end{aligned}$$

where

$$W_n = 1 - 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) . \quad (3.63)$$

Thus we have

$$\begin{aligned}
x_{1,n+1} &= \frac{1}{S} \{ (1-r)x_{1,n}x_{4,n} + rx_{2,n}x_{3,n} + x_{1,n}^2 (1 - \frac{1-S}{W_n}) \\
& + x_{1,n}x_{3,n} (1 - \frac{1-S}{W_n}) + x_{1,n}x_{2,n} (1 - \frac{1-S}{W_n}) \} \\
&= \frac{1}{S} [x_{1,n} - rD_n - x_{1,n} \frac{(1-S)}{W_n} (1-x_{4,n})] \\
x_{2,n+1} &= \frac{1}{S} \{ (1-r)x_{2,n}x_{3,n} + rx_{1,n}x_{4,n} + x_{1,n}x_{2,n} (1 - \frac{1-S}{W_n}) \\
& + x_{2,n}^2 (1 - \frac{1-S}{W_n}) + x_{2,n}x_{4,n} (1 - \frac{1-S}{W_n}) \} \\
&= \frac{1}{S} [rD_n + x_{2,n} - \frac{(1-S)}{W_n} x_{2,n}(1-x_{3,n})]
\end{aligned} \quad (3.64)$$

$$\begin{aligned}
x_{3,n+1} &= \frac{1}{S} \{ (1-r)x_{2,n}x_{3,n} + rx_{1,n}x_{4,n} + x_{1,n}x_{3,n}(1 - \frac{1-S}{W_n}) \\
&\quad + x_{3,n}^2 (1 - \frac{1-S}{W_n}) + x_{3,n}x_{4,n}(1 - \frac{1-S}{W_n}) \} \\
&= \frac{1}{S} [rD_n + x_{3,n} - \frac{1-S}{W_n} x_{3,n}(1-x_{2,n})] \\
x_{4,n+1} &= \frac{1}{S} \{ (1-r)x_{1,n}x_{4,n} + rx_{2,n}x_{3,n} + x_{3,n}x_{4,n}(1 - \frac{1-S}{W_n}) \\
&\quad + x_{2,n}x_{4,n} (1 - \frac{1-S}{W_n}) + x_{4,n}^2 (1 - \frac{1-S}{W_n}) \} \\
&= \frac{1}{S} [x_{4,n} - rD_n - \frac{1-S}{W_n} x_{4,n}(1-x_{1,n})] ,
\end{aligned}$$

where W_n is given by (3.63). Equations (3.64) show that linkage may have some effect on the changes in the gametic frequencies. We now consider the differences $(x_{1,n+1} - x_{4,n+1})$ and $(x_{3,n+1} - x_{2,n+1})$. Using (3.64), we find

$$\begin{aligned}
x_{1,n+1} - x_{4,n+1} &= \frac{1}{S} (x_{1,n} - x_{4,n})(1 - \frac{1-S}{W_n}) \\
x_{3,n+1} - x_{2,n+1} &= \frac{1}{S} (x_{3,n} - x_{2,n})(1 - \frac{1-S}{W_n})
\end{aligned}$$

where W_n is defined in (3.63). So, as long as the inequality on S holds,

$$\begin{aligned}
x_{1,n} - x_{4,n} &= \frac{1}{S^n} (x_{1,o} - x_{4,o}) \prod_{i=0}^{n-1} (1 - \frac{1-S}{W_i}) \\
x_{3,n} - x_{2,n} &= \frac{1}{S^n} (x_{3,o} - x_{4,o}) \prod_{i=0}^{n-1} (1 - \frac{1-S}{W_i}) .
\end{aligned} \tag{3.65}$$

Equations (3.65) show that if $x_{1,0} = x_{4,0}$ and $x_{2,0} = x_{3,0}$, we have in every generation $x_1 = x_4 = \frac{1}{2}$ - $x_2 = \frac{1}{2}$ - x_3 . If $x_{1,0} \neq x_{4,0}$ and $x_{2,0} \neq x_{3,0}$, the differences $(x_1 - x_4)$ and $(x_3 - x_2)$ will be increased by selection. We note here that, from the defining inequality for this stage, the term $(1 - \frac{1-S}{W_i})$ appearing in (3.65) is greater than zero.

The new frequencies of the A and B genes are given by

$$p_{n+1} = x_{1,n+1} + x_{2,n+1}$$

$$q_{n+1} = x_{1,n+1} + x_{3,n+1} .$$

Using (3.64) we find

$$\begin{aligned} p_{n+1} &= \frac{1}{S} p_n \left(1 - \frac{1-S}{W_n}\right) + \frac{1-S}{SW_n} (x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) \\ q_{n+1} &= \frac{1}{S} q_n \left(1 - \frac{1-S}{W_n}\right) + \frac{1-S}{SW_n} (x_{1,n}x_{4,n} + x_{2,n}x_{3,n}). \end{aligned} \quad (3.66)$$

The difference between the frequencies of the two genes in generation n is given by $(x_{3,n} - x_{2,n})$ which using (3.65) can be written in the form

$$q_n - p_n = \frac{1}{S^n} (q_0 - p_0) \prod_{i=0}^{n-1} \left(1 - \frac{1-S}{W_i}\right) , \quad (3.67)$$

where W_i is defined in (3.63).

The gametic disequilibrium in the (n+1)th generation is given by

$$D_{n+1} = x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1} .$$

Using (3.64), we find

$$\begin{aligned}
 D_{n+1} = & \frac{1}{S^2} \left\{ 1 - r \left[1 - \frac{1-S}{W_n} + \frac{1-S}{W_n} (1 + x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) \left(2 + \frac{1-S}{W_n} \right) \right] \right. \\
 & + \left. \left(\frac{1-S}{W_n} \right) \left(1 - \frac{1-S}{W_n} \right) (1 + x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) \right\} D_n \\
 & - \frac{1}{S^2} \left(\frac{1-S}{W_n} \right) \left(1 - \frac{1-S}{W_n} \right) [x_{1,n}x_{2,n}(x_{4,n} - x_{3,n}) + x_{3,n}x_{4,n}(x_{1,n} - x_{2,n})].
 \end{aligned} \quad (3.68)$$

For the symmetrical case $x_1 = x_4 = \frac{1}{2} - x_2 = \frac{1}{2} - x_3$, equations (3.64) can be written in the form

$$x_{i,n+1} = \frac{1}{S} \{ \pm r D_n + (1 - \frac{1-S}{W_n}) x_{i,n} + \frac{1-S}{W_n} x_{i,n}^2 \}, \quad i = 1, 2, 3, 4 \quad (3.69)$$

where the sign of $(r D_n)$ is negative for $i = 1$ and 4 and is positive for $i = 2$ and 3 ,

$$W_n = \frac{1}{2} + 2x_{i,n} - 4x_{i,n}^2,$$

and

$$D_n = x_{1,n} - \frac{1}{4} = \frac{1}{4} - x_{2,n}.$$

Using (3.69), it is seen that $p_n = q_n = \frac{1}{2}$. The change in the gametic frequencies is given by

$$\Delta x_{i,n} = x_{i,n+1} - x_{i,n}.$$

Using (3.69) we find

$$\begin{aligned}
 \Delta x_{1,n} &= \frac{1}{S} \left[-rD_n + \left(1-S - \frac{1-S}{W_n}\right) x_{1,n} + \frac{1-S}{W_n} x_{1,n}x_{4,n} \right] \\
 \Delta x_{2,n} &= \frac{1}{S} \left[rD_n + \left(1-S - \frac{1-S}{W_n}\right) x_{2,n} + \frac{1-S}{W_n} x_{2,n}x_{3,n} \right] \\
 \Delta x_{3,n} &= \frac{1}{S} \left[rD_n + \left(1-S - \frac{1-S}{W_n}\right) x_{3,n} + \frac{1-S}{W_n} x_{2,n}x_{3,n} \right] \\
 \Delta x_{4,n} &= \frac{1}{S} \left[-rD_n + \left(1-S - \frac{1-S}{W_n}\right) x_{4,n} + \frac{1-S}{W_n} x_{1,n}x_{4,n} \right] .
 \end{aligned} \tag{3.70}$$

For the symmetrical case, $x_1 = x_4 = \frac{1}{2} - x_2 = \frac{1}{2} - x_3$, we have

$$\Delta x_{1,n} = \frac{\frac{r}{8} - \frac{1}{2}(1-S)x_{1,n} + 3(1-r-S)x_{1,n}^2 - 4(1-r-S)x_{1,n}^3}{S\left(\frac{1}{2} + 2x_{1,n} - 4x_{1,n}^2\right)} \tag{3.71}$$

Equilibrium exists when $\Delta x_{i,n} = 0$ for all i . For the symmetrical case, $x_1 = x_4$ and $x_2 = x_3$, the equilibrium frequencies are given by the roots of the cubic equation

$$\frac{r}{8} - \frac{1}{2}(1-S)x_1 + 3(1-r-S)x_{1,n}^2 - 4(1-r-S)x_{1,n}^3 = 0 \tag{3.72}$$

The roots are functions of S and r . It is seen from (3.72) that when $r = \frac{1}{2}$ the three roots are all equal to $\frac{1}{4}$.

$$2. \quad (1-S) \geq 1 - 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n}).$$

In this stage no genotypes except the double heterozygotes $AaBb$ survive to breed. Thus we find

$$\begin{aligned}
 x_{1,n+1} &= x_{4,n+1} = \frac{1}{W_n} (x_{1,n}x_{4,n} - rD_n) \\
 x_{2,n+1} &= x_{3,n+1} = \frac{1}{W_n} (x_{2,n}x_{3,n} + rD_n) ,
 \end{aligned}
 \tag{3.73}$$

where

$$W_n = 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) . \tag{3.74}$$

Equations (3.73) show that linkage may have some effect on the changes in the gametic frequencies. Linkage will accelerate the rate of change in the frequency of the gametes AB and ab when $D > 0$ and retard the rate of change when $D < 0$. Equations (3.73) imply that

$$x_1 = x_4 = \frac{1}{2} - x_2 = \frac{1}{2} - x_3 \tag{3.75}$$

in any generation during the second stage. Consequently, the gametic disequilibrium in generation n is given by

$$\begin{aligned}
 D_n &= x_{1,n}x_{4,n} - x_{2,n}x_{3,n} \\
 &= x_{1,n} - \frac{1}{4} \\
 &= \frac{1}{4} - x_{2,n} .
 \end{aligned}
 \tag{3.76}$$

Replacing D_n in (3.73) by its value from (3.76) we find

$$x_{i,n+1} = \frac{\frac{r}{4} - rx_{i,n} + x_{i,n}^2}{\frac{1}{2} - 2x_{i,n} + 4x_{i,n}^2} \quad i = 1, 2, 3, 4 . \tag{3.77}$$

Consider now the difference

$$x_{1,n+1} - x_{2,n+1} = \frac{1}{W_n} (1-2r)D_n , \tag{3.78}$$

where W_n is defined in (3.74). Equation (3.78) shows that, when $r = \frac{1}{2}$, we have $x_1 = x_2 = x_3 = x_4 = \frac{1}{4}$ in every generation.

The new frequencies of the A and B genes are given by

$$p_{n+1} = x_{1,n+1} + x_{2,n+1}$$

$$q_{n+1} = x_{1,n+1} + x_{3,n+1} ,$$

respectively. Using (3.73), we find

$$p_{n+1} = q_{n+1} = \frac{1}{2} \quad (3.79)$$

The gametic disequilibrium in the (n+1)th generation is given by

$$D_{n+1} = x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1} .$$

Using (3.73), we find

$$D_{n+1} = \frac{1}{4} \frac{1-2r}{(x_{1,n}x_{4,n} + x_{2,n}x_{3,n})} D_n . \quad (3.80)$$

Equation (3.80) for the change in gametic disequilibrium shows that

(i) gametic equilibrium will be maintained in the population, i.e. $D = 0$, when there is no linkage, $r = \frac{1}{2}$, whether D_0 is equal or unequal to zero,

(ii) when $D_0 = 0$, the population will be in gametic equilibrium irrespective of linkage, and

(iii) when $D_0 \neq 0$, the tighter the linkage the longer the population will be in gametic disequilibrium.

Figure 5 shows the change in D with generations under different linkage intensities when $D_0 = -.2$. The figure is based on the exact

solutions which are in agreement with equation (3.80) above. Equation (3.80) for the gametic disequilibrium is similar to that derived by Haldane (1962) when recombination is the same in the two sexes, with $t = 2D$ and $W = 1$. He pointed out that, when linkage is loose or selection is weak, coupling and repulsion should be equally frequent and $D = 0$. These observations are in agreement with equations (3.75) and (3.80). The change in gametic disequilibrium in one generation is given by

$$\begin{aligned} D_n &= D_{n+1} - D_n \\ &= (x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1}) - (x_{1,n}x_{4,n} - x_{2,n}x_{3,n}) \\ &= \frac{1}{W_n} \left(\frac{1}{2} - r - W_n \right) D_n, \end{aligned} \quad (3.81)$$

where W_n is defined in (3.74). Using (3.75), we find

$$\Delta D_n = \Delta x_{1,n}$$

or approximately,

$$\frac{dD}{dn} = \frac{dx_1}{dn}. \quad (3.82)$$

Equation (3.82) indicates that the conditions for an equilibrium in the gametic frequencies are the same as those required for gametic equilibrium or disequilibrium to be constant.

Using equation (3.77), the change in the gametic frequencies after one generation may be written in the form

$$\Delta x_{i,n} = W_i^{-1} \left(\frac{1}{2} r D_n - \frac{1}{2} x_{i,n} + 3x_{i,n}^2 - 4x_{i,n}^3 \right), \quad (3.83)$$

where

$$W_i = \frac{1}{2} - 2x_{i,n} + 4x_{i,n}^2 ,$$

with $i = 1, 2, 3, 4$, and the sign of (rD_n) is negative for $i = 1$ and 4 and positive for $i = 2$ and 3 . Equilibrium in gametic frequencies exist when $\Delta x_{i,n} = 0$ for all i . Because of the symmetry under the present model, the solution requires that

$$x_1 = x_4 = \frac{1}{2} - x_2 = \frac{1}{2} - x_3$$

and

$$D = x_1 - \frac{1}{4} ,$$

at equilibrium. Because equations (3.83) are cubic in x_1 , there are three roots, each gives rise to an equilibrium point. When $D_0 = 0$ or $r = \frac{1}{2}$, it is seen that the roots are 0 , $\frac{1}{4}$ and $\frac{1}{2}$. When $D_0 \neq 0$ and $0 \leq r < \frac{1}{2}$, the equilibrium points are those which satisfy the equation

$$x_1^3 - \frac{3}{4} x_1^2 + \frac{1}{4} \left(\frac{1}{2} + r \right) x_1 - \frac{r}{16} = 0 .$$

When $r = \frac{1}{4}$, there are three roots and all equal to $\frac{1}{4}$. When $r > \frac{1}{4}$ there will be one real root and two conjugate imaginary roots. The real root is given by $x = \frac{1}{4}$. When $r < \frac{1}{4}$, there will be three real and unequal roots. These roots are given in Table 17 for different values of r .

Table 17. Equilibrium points for different values of linkage r .

Roots for equilibrium frequency x_1	Linkage r				
	0	.10	.15	.20	.25
First root	.250	.250	.250	.250	.250
Second root	.253	.2514	.2507	.2502	.250
Third root	.247	.2496	.2493	.2498	.25

We consider now the stability of the above equilibrium. Bodmer and Felsenstein (1967) examined the stability by considering the ratio

$|D_{n+1}|/|D_n|$. The equilibrium is said to be stable if the ratio is less than one in absolute value. This is equivalent to an equilibrium taking place in the neighborhood of $D = 0$. From equation (3.80) we have

$$\frac{D_{n+1}}{D_n} = W^{-1} \left(\frac{1}{2} - r \right),$$

where W is defined in (3.74). Thus in order for the absolute ratio to be less than one we must have $r > \frac{1}{2} - (x_1x_4 + x_2x_3)$. Hence for equilibrium point of $\frac{1}{4}$ to be stable we must have $r > \frac{1}{4}$. This is in agreement with the findings by Bodmer and Felsenstein if we put $t_1 = t_2 = s_1 = s_2 = 1$ in their notations, where t_i and s_i are used to define the relative viability of genotypes and which are assumed to be constant (Table 18).

Table 18. Relative viability of genotypes (Bodmer and Felsenstein, 1967).

	AA	Aa	aa
BB	$(1-s_1)(1-s_2)$	$(1-s_2)$	$(1-t_1)(1-s_2)$
Bb	$(1-s_1)$	1	$(1-t_1)$
bb	$(1-s_1)(1-t_2)$	$(1-t_2)$	$(1-t_1)(1-t_2)$

They pointed out that two multiplicative over-dominant loci cannot be at a stable equilibrium with $D = 0$ if the recombination fraction between them is less than the product of their segregational loads, i.e.

$$r > \left(\frac{s_1 t_1}{s_1 + t_1} \right) \left(\frac{s_2 t_2}{s_2 + t_2} \right) .$$

In the present study, it is seen that this product is $\frac{1}{4}$. If $r < \frac{1}{4}$, one or more other points different from $\frac{1}{4}$ with $D \neq 0$ will be stable.

We consider now a different approach for examining the stability of the equilibria obtained above. It is clear that when $D = 0$, an equilibrium point of $\frac{1}{4}$ is expected and it is stable. In order to examine the stability of the equilibrium when $D \neq 0$, we let x_1, x_2, x_3 and x_4 be the equilibrium frequencies of the four gametes, and suppose that

$$\begin{aligned} x_{1,n} &= x_1 + \delta_{1,n} \\ x_{2,n} &= x_2 + \delta_{2,n} \\ x_{3,n} &= x_3 + \delta_{3,n} \\ x_{4,n} &= x_4 + \delta_{4,n} , \end{aligned} \tag{3.84}$$

where $\sum \delta_{i,n} = 0$ and $\delta_{i,n}$ are small for $i = 1, 2, 3, 4$. From (3.73) we have at equilibrium

$$\begin{aligned} x_1 &= (x_1 x_4 - rD) / 2(x_1 x_4 + x_2 x_3) \\ x_2 &= (x_2 x_3 + rD) / 2(x_1 x_4 + x_2 x_3) \\ x_3 &= (x_2 x_3 + rD) / 2(x_1 x_4 + x_2 x_3) \\ x_4 &= (x_1 x_4 - rD) / 2(x_1 x_4 + x_2 x_3) , \end{aligned} \quad (3.85)$$

where D is the gametic disequilibrium. From (3.73), (3.84) and (3.85) the $\delta_{i,n+1}$ can be expressed as linear functions of the $\delta_{i,n}$. Substituting for $x_{i,n}$ in (3.73) by their values from (3.84) we get

$$\begin{aligned} x_{1,n+1} &= \frac{x_1 x_4 - rD}{2(x_1 x_4 + x_2 x_3)} - \frac{(1-r)x_4 \delta_{1,n} + (1-r)x_1 \delta_{4,n} + rx_2 \delta_{3,n} + rx_3 \delta_{2,n}}{2(x_1 x_4 + x_2 x_3)} \\ &= x_1 + \delta_{1,n+1} . \end{aligned}$$

We now consider the stability of the equilibrium given by $x_1 = x_2 = x_3 = x_4 = \frac{1}{4}$. For this equilibrium point the matrix of the $\delta_{i,n+1}$ is given approximately by

$$\begin{aligned} \delta_{n+1} &= \begin{bmatrix} \delta_{1,n+1} \\ \delta_{2,n+1} \\ \delta_{3,n+1} \\ \delta_{4,n+1} \end{bmatrix} = \begin{bmatrix} (1-r) & r & r & (1-r) \\ r & (1-r) & (1-r) & r \\ r & (1-r) & (1-r) & r \\ (1-r) & r & r & (1-r) \end{bmatrix} \begin{bmatrix} \delta_{1,n} \\ \delta_{2,n} \\ \delta_{3,n} \\ \delta_{4,n} \end{bmatrix} \\ &= V \delta_n . \end{aligned} \quad (3.86)$$

The equilibrium given by (x_1, x_2, x_3, x_4) will therefore be stable if and only if all the characteristic roots of the matrix V are less than one in absolute value. When this is the case, δ_n , the vector of departure from equilibrium will converge to a zero vector as n increases. The characteristic roots, λ_i , are obtained from the equations

$$0 = |V - \lambda I| ,$$

where I is the identity matrix. It is clear from (3.86) that zero is one of the roots. The other roots are to be obtained from the quadratic equation

$$\lambda^2 - 4(1-r)\lambda + 4(1-2r) = 0 . \quad (3.87)$$

One of the roots depends on linkage and is given by $\lambda = 2(1-2r)$. Thus for the equilibrium point of $\frac{1}{4}$ to be stable we must have $r > \frac{1}{4}$. This is in agreement with the findings by Lewontin and Kojima (1960). They examined the equilibrium conditions for a population subjected to natural selection using a symmetric system with relative viabilities of genotypes assumed constant, and with the assumptions that there is no difference in fecundity and no environmental variability. Their model is similar to the one presented in this section (Table 27, Appendix B) if we put $a = b = c = 0$ and $d > 0$, a condition which amounts to over-dominance of each locus with interaction. It is seen from their findings that, when $D = 0$, there is a stable equilibrium in the gametic frequencies given by $x_1 = x_2 = x_3 = x_4 = \frac{1}{4}$. When $D \neq 0$, a stable equilibrium of $\frac{1}{4}$ is attained if $r > \frac{1}{4}$, and two other equilibrium points different from $\frac{1}{4}$ which will be stable if $r \leq \frac{1}{4}$.

We now consider the stability of the equilibrium given by $x_1 = x_4 = \frac{1}{2}$ and $x_2 = x_3 = 0$. The matrix $(V - \lambda I)$ is given by

$$(V - \lambda I) = \begin{bmatrix} (1-r)-\lambda & 0 & 0 & (1-r) \\ r & -\lambda & 0 & r \\ r & 0 & -\lambda & r \\ (1-r) & 0 & 0 & (1-r)-\lambda \end{bmatrix} .$$

The characteristic roots λ_i are obtained from the equation

$$\lambda^4 - 2(1-r)\lambda^3 = 0 ,$$

which is the same equation for the equilibrium given by $x_1 = x_4 = 0$ and $x_2 = x_3 = \frac{1}{2}$. The roots are therefore $\lambda = 0$ and $\lambda = 2(1-r)$. It is clear that the root which depends on linkage cannot be less than one in absolute value. Thus an equilibrium given by $x_1 = x_4 = \frac{1}{2}$ and $x_2 = x_3 = 0$ is unstable. This indicates that if there are stable equilibrium points other than $\frac{1}{4}$, these points must be close to $\frac{1}{4}$ (Table 17). Parsons (1963) discussed the equilibrium conditions for population subjected to natural selection using a symmetric model with the assumptions that the relative viabilities of genotypes are constant, no environmental variability and no difference in fecundity. The model he considered is presented in Table 19.

Table 19. Relative viabilities of genotypes (Parsons, 1963).

	AA	Aa	aa
BB	1- γ	1- β	1
Bb	1- β	1+ α	1- β
bb	1	1- β	1- γ

Thus his model is similar to our model if we put $\gamma = \beta = 0$ and $\alpha > 0$. He pointed out that when $r > \frac{1}{4}$, there is only one stable equilibrium given by $\frac{1}{4}$. When $r < \frac{1}{4}$, there will be three equilibrium points, one being $\frac{1}{4}$ and the other two are different but close to $\frac{1}{4}$.

In order to continuize the change in gametic frequencies under the present symmetric model we write

$$\Delta x_{i,n} = x_{i,n+1} - x_{i,n} \doteq \frac{dx_i}{dn}.$$

Using equation (3.83) and assuming that either D or r is zero, we find

$$n = \ln \frac{(1 - 4x_{1,n})}{(1 - 2x_{1,n})} - \ln x_{1,n} - \ln \frac{(1 - 4x_{1,o})}{(1 - 2x_{1,o})} + \ln x_{1,o}. \quad (3.88)$$

The above function of $x_{1,n}$ is expected to change linearly with generations provided that either D or r is zero. When $D \neq 0$, a general solution can be obtained by solving the differential equation

$$dn = \frac{\frac{1}{2} - 2x_1 + 4x_1^2}{\frac{1}{4}r - (\frac{1}{2} + r)x_1 + 3x_1^2 - 4x_1^3} dx_1. \quad (3.89)$$

The solution of the above differential equation is given by

$$n = \frac{1}{3} (1-r) \left\{ \frac{\ln(x_1 - u_1)}{-12u_1^2 + 6u_1 - (\frac{1}{2} + r)} + \frac{\ln(x_1 - u_2)}{-12u_2^2 + 6u_2 - (\frac{1}{2} + r)} \right. \\ \left. + \frac{\ln(x_1 - u_3)}{-12u_3^2 + 6u_3 - (\frac{1}{2} + r)} - \frac{1}{3} \ln \left[\frac{1}{4}r - (\frac{1}{2} + r)x_1 + 3x_1^2 - x_1^3 \right] \right\} \frac{x_{1,n}}{x_{1,o}} \quad (3.90)$$

where u_i are the roots of the cubic equation

$$\frac{1}{4}r - (\frac{1}{2} + r)x_1 + 3x_1^2 - 4x_1^3 = 0.$$

The roots are given in Table 17 for different values of linkage. It is seen from (3.90) that the slope of the linear function of the gametic frequency is given by $\hat{\beta} = 3/(1-r)$ and therefore decreases with linkage.

Overdominance without interaction

We now consider an overdominance model with cumulative heterozygous advantage and without epistatic effects. The mean phenotypic values for the present model are presented in Table 27, Appendix B. The order of the genotypes is $aabb = aaBB = AAbb = AABB < Aabb = AABb = aaBb = AaBB < AB/ab = Ab/aB$. There are three possible stages of progress.

$$1. \quad (1-S) < x_{4,n}^2 + x_{3,n}^2 + x_{2,n}^2 + x_{1,n}^2.$$

In this stage there are some $aabb$, $AAbb$, $aaBB$ and $AABB$ among the individuals saved. The array among the culled individuals in generation n is

$$\frac{1}{\sum x_{i,n}^2} \{x_{4,n}^2 aabb + x_{3,n}^2 aaBB + x_{2,n}^2 AAbb + x_{1,n}^2 AABB\}.$$

The array among the saved individuals in the same generation is

$$\begin{aligned} & \frac{1}{S} \{2x_{1,n}x_{4,n} AB/ab + 2x_{2,n}x_{3,n} Ab/aB + 2x_{1,n}x_{3,n} AaBB + 2x_{3,n}x_{4,n} aaBb \\ & + 2x_{1,n}x_{2,n} AABb + 2x_{2,n}x_{4,n} Aabb + x_{1,n}^2 (1 - \frac{1-S}{\sum x_{i,n}^2}) AABB \\ & + x_{2,n}^2 (1 - \frac{1-S}{\sum x_{i,n}^2}) AAbb + x_{3,n}^2 (1 - \frac{1-S}{\sum x_{i,n}^2}) aaBB \\ & + x_{4,n}^2 (1 - \frac{1-S}{\sum x_{i,n}^2}) aabb\}. \end{aligned}$$

Thus we have

$$\begin{aligned}
 x_{1,n+1} &= \frac{1}{S} \{ (1-r) x_{1,n} x_{4,n} + r x_{2,n} x_{3,n} + x_{1,n} x_{2,n} + x_{1,n} x_{3,n} \\
 &\quad + x_{1,n}^2 - \frac{1}{\sum x_{i,n}^2} (1-S) x_{1,n}^2 \} \\
 &= \frac{1}{S} \left[x_{1,n} - rD_n - \frac{1}{\sum x_{i,n}^2} (1-S) x_{1,n}^2 \right]
 \end{aligned}$$

$$\begin{aligned}
 x_{2,n+1} &= \frac{1}{S} \{ r x_{1,n} x_{4,n} + (1-r) x_{2,n} x_{3,n} + x_{1,n} x_{2,n} + x_{2,n} x_{4,n} \\
 &\quad + x_{2,n}^2 - \frac{1}{\sum x_{i,n}^2} (1-S) x_{2,n}^2 \} \\
 &= \frac{1}{S} \left[x_{2,n} + rD_n - \frac{1}{\sum x_{i,n}^2} (1-S) x_{2,n}^2 \right]
 \end{aligned}$$

(3.91)

$$\begin{aligned}
 x_{3,n+1} &= \frac{1}{S} \{ r x_{1,n} x_{4,n} + (1-r) x_{2,n} x_{3,n} + x_{1,n} x_{3,n} + x_{3,n} x_{4,n} \\
 &\quad + x_{3,n}^2 - \frac{1}{\sum x_{i,n}^2} (1-S) x_{3,n}^2 \} \\
 &= \frac{1}{S} \left[x_{3,n} + rD_n - \frac{1}{\sum x_{i,n}^2} (1-S) x_{3,n}^2 \right]
 \end{aligned}$$

$$\begin{aligned}
 x_{4,n+1} &= \frac{1}{S} \{ (1-r) x_{1,n} x_{4,n} + r x_{2,n} x_{3,n} + x_{2,n} x_{4,n} + x_{3,n} x_{4,n} \\
 &\quad + x_{4,n}^2 - \frac{1}{\sum x_{i,n}^2} (1-S) x_{4,n}^2 \} \\
 &= \frac{1}{S} \left[x_{4,n} - rD_n - \frac{1}{\sum x_{i,n}^2} (1-S) x_{4,n}^2 \right] .
 \end{aligned}$$

The gametic disequilibrium in the (n+1)th generation is given by

$$D_{n+1} = x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1} .$$

Using (3.91), we find

$$\begin{aligned} D_{n+1} = & \frac{1}{S^2} \left[1-r S + \left(\frac{1-S}{\sum x_{i,n}^2} \right)^2 (x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) \right] D_n \\ & - \frac{1-S}{S^2 \sum x_{i,n}^2} \left[(x_{1,n}x_{4,n} + x_{2,n}x_{3,n})(x_{1,n} + x_{4,n}) - x_{2,n}x_{3,n} \right] . \end{aligned} \quad (3.92)$$

Equation (3.92) shows that selection may build up gametic disequilibrium even when the initial population is in gametic equilibrium, i.e. $D_0 = 0$.

The new frequencies of the A and B genes are given by

$$\begin{aligned} p_{n+1} &= x_{1,n+1} + x_{2,n+1} \\ q_{n+1} &= x_{1,n+1} + x_{3,n+1} , \end{aligned}$$

respectively. Using (3.91), we find

$$\begin{aligned} p_{n+1} &= \frac{1}{S} p_n - \frac{1-S}{S \sum x_{i,n}^2} (x_{1,n}^2 + x_{2,n}^2) \\ q_{n+1} &= \frac{1}{S} q_n - \frac{1-S}{S \sum x_{i,n}^2} (x_{1,n}^2 + x_{3,n}^2) . \end{aligned} \quad (3.93)$$

We now consider the difference

$$q_{n+1} - p_{n+1} = \frac{1}{S} (q_n - p_n) \left[1 - \frac{1-S}{\sum x_{i,n}^2} (x_{3,n} + x_{2,n}) \right] .$$

So, as long as the inequality on S holds,

$$q_n - p_n = \frac{1}{s^n} (q_0 - p_0) \prod_{j=0}^{n-1} \left[1 - \frac{1-s}{\sum x_{i,j}^2} (x_{3,j} + x_{2,j}) \right]. \quad (3.94)$$

Equation (3.94) shows that when $q_0 = p_0$ or $x_{2,0} = x_{3,0}$ the frequencies of the two genes A and B will remain equal in every generation. When $q_0 \neq p_0$, the difference $(q_n - p_n)$ will be increased by selection. Note here that the term $1 - \frac{1-s}{\sum x_{i,j}^2} (x_{2,j} + x_{3,j})$ appearing in (3.94) is greater than zero.

Consider now the symmetrical case, $x_1 = x_4 = \frac{1}{2} - x_2 = \frac{1}{2} - x_3$. For this case, equations (3.92), (3.93) and (3.94) reduce to

$$D_{n+1} = \frac{1}{s^2} \left[1-r-s + \frac{(1-s)^2}{2 \sum x_{i,n}^2} \right] D_n - \frac{1-s}{s^2} \left[x_{1,n} - \frac{(\frac{1}{2} - x_{1,n})^2}{\sum x_{i,n}^2} \right] \quad (3.95)$$

$$p_{n+1} = \frac{1}{s} p_n - \frac{1-s}{2s} \quad (3.96)$$

$$q_{n+1} = \frac{1}{s} q_n - \frac{1-s}{2s}$$

$$q_n - p_n = \frac{1}{s^n} (q_0 - p_0), \quad (3.97)$$

respectively. We also have

$$x_{1,n+1} = \frac{\frac{r}{8} + (\frac{1}{2} - r)x_{1,n} - (3-3r-s)x_{1,n}^2 + 4(1-r)x_{1,n}^3}{s(\frac{1}{2} - 2x_{1,n} + 4x_{1,n}^2)}. \quad (3.98)$$

The change in the gametic frequencies is given by

$$\Delta x_{i,n} = x_{i,n+1} - x_{i,n} \quad i = 1, 2, 3, 4 .$$

Using (3.91), we find

$$\Delta x_{i,n} = \frac{1}{S} \left\{ (1-S)x_{i,n} - \frac{1}{\sum x_{i,n}^2} (1-S) x_{i,n}^2 \pm rD_n \right\} \quad (3.99)$$

where the sign of (rD_n) is positive for $i = 2$ and 3 and is negative for $i = 1$ and 4 . For the symmetrical case $x_1 = x_4$ and $x_2 = x_3$, equation (3.99) may be written in the form

$$\Delta x_{1,n} = \frac{\frac{r}{8} + \frac{1}{2} (1-2r-S)x_{1,n} - 3(1-r-S)x_{1,n}^2 + 4(1-r-S)x_{1,n}^3}{S(\frac{1}{2} - 2x_{1,n} + 4x_{1,n}^2)} . \quad (3.100)$$

Equilibrium exists when $\Delta x_{i,n} = 0$ for all i . For the symmetrical case, it is seen from (3.100) that equilibrium exists when $x_1 = 0, \frac{1}{4}, \frac{1}{2}$.

$$2. \quad x_{4,n}^2 + x_{3,n}^2 + x_{2,n}^2 + x_{1,n}^2 \leq 1-S < 1 - 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) .$$

In this stage all of the $aabb$, $aaBB$, $Aabb$, $AABB$ and some of the $Aabb$, $AABb$, $aaBb$ and $AaBB$ individuals are culled. The array among the culled individuals in generation n is

$$\frac{2x_{2,n}x_{4,n} Aabb + 2x_{1,n}x_{2,n} AABb + 2x_{3,n}x_{4,n} aaBb + 2x_{1,n}x_{3,n} AaBB}{2(x_{2,n}x_{4,n} + x_{1,n}x_{2,n} + x_{3,n}x_{4,n} + x_{1,n}x_{3,n})} .$$

The array among the saved individuals in the same generation is

$$\begin{aligned}
& \frac{1}{S} \{ 2x_{2,n}x_{3,n} \text{ Ab/aB} + 2x_{1,n}x_{4,n} \text{ AB/ab} \\
& + 2x_{2,n}x_{4,n} [1-(1-S-\Sigma x_{i,n}^2) V^{-1}] \text{ Aabb} \\
& + 2x_{1,n}x_{2,n} [1-(1-S-\Sigma x_{i,n}^2) V^{-1}] \text{ AABb} \\
& + 2x_{3,n}x_{4,n} [1-(1-S-\Sigma x_{i,n}^2) V^{-1}] \text{ aaBb} \\
& + 2x_{1,n}x_{3,n} [1-(1-S-\Sigma x_{i,n}^2) V^{-1}] \text{ AaBB} \} ,
\end{aligned}$$

where

$$V = 2(x_{2,n}x_{4,n} + x_{1,n}x_{2,n} + x_{3,n}x_{4,n} + x_{1,n}x_{3,n}) . \quad (3.101)$$

Thus we have

$$\begin{aligned}
x_{1,n+1} &= \frac{1}{S} \{ r x_{2,n}x_{3,n} + (1-r)x_{1,n}x_{4,n} + x_{1,n}x_{2,n} [1-(1-S-\Sigma x_{i,n}^2) V^{-1}] \\
& + x_{1,n}x_{3,n} [1-(1-S-\Sigma x_{i,n}^2) V^{-1}] \} \\
&= \frac{1}{S} [x_{1,n}(1-x_{1,n}) - rD_n - x_{1,n}(x_{2,n} + x_{3,n})(1-S-\Sigma x_{i,n}^2) V^{-1}] \\
x_{2,n+1} &= \frac{1}{S} \{ (1-r)x_{2,n}x_{3,n} + r x_{1,n}x_{4,n} + x_{2,n}x_{4,n} [1-(1-S-\Sigma x_{i,n}^2) V^{-1}] \\
& + x_{1,n}x_{2,n} [1-(1-S-\Sigma x_{i,n}^2) V^{-1}] \} \\
&= \frac{1}{S} [x_{2,n}(1-x_{2,n}) + rD_n + x_{2,n}(x_{1,n} + x_{4,n})(1-S-\Sigma x_{i,n}^2) V^{-1}] \\
& \quad (3.102)
\end{aligned}$$

$$\begin{aligned}
x_{3,n+1} &= \frac{1}{S} \{ (1-r)x_{2,n}x_{3,n} + r x_{1,n}x_{4,n} + x_{3,n}x_{4,n} [1-(1-S-\Sigma x_{i,n}^2) V^{-1}] \\
& + x_{1,n}x_{3,n} [1-(1-S-\Sigma x_{i,n}^2) V^{-1}] \}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{S} \{x_{3,n}(1-x_{3,n}) + rD_n - x_{3,n}(x_{1,n} + x_{4,n})(1-S-\Sigma x_{i,n}^2) V^{-1}\} \\
x_{4,n+1} &= \frac{1}{S} \{r x_{2,n}x_{3,n} + (1-r)x_{1,n}x_{4,n} + x_{2,n}x_{4,n} [1-(1-S-\Sigma x_{i,n}^2) V^{-1}] \\
&\quad + x_{3,n}x_{4,n} [1-(1-S-\Sigma x_{i,n}^2) V^{-1}]\} \\
&= \frac{1}{S} \{x_{4,n}(1-x_{4,n}) - rD_n - x_{4,n}(x_{2,n} + x_{3,n})(1-S-\Sigma x_{i,n}^2) V^{-1}\} ,
\end{aligned}$$

where V is defined in (3.101). Because of the symmetry of equations (3.102) it is seen that when $x_{1,o} = x_{4,o}$ and $x_{2,o} = x_{3,o}$ we have in every generation $x_1 = x_4 = \frac{1}{2} - x_2 = \frac{1}{2} - x_3$. For this symmetrical case, we have $D_n = x_{1,n} - \frac{1}{4}$, and equations (3.102) reduce to

$$x_{i,n+1} = \frac{1}{S} \left[\left(\frac{1}{2} - r \right) x_{i,n} - \frac{1}{4} \left(\frac{1}{2} - r - S \right) \right] \quad i = 1, 2, 3, 4. \quad (3.103)$$

Equation (3.103) shows that, for the symmetrical case $x_1 = x_4$ and $x_2 = x_3$ and when $r = \frac{1}{2}$, we have in every generation $x_1 = x_2 = x_3 = x_4 = \frac{1}{4}$.

The new frequencies of the A and B genes are given by

$$\begin{aligned}
p_{n+1} &= x_{1,n+1} + x_{2,n+1} \\
q_{n+1} &= x_{1,n+1} + x_{3,n+1} .
\end{aligned}$$

Using (3.102), we find

$$\begin{aligned}
p_{n+1} &= \frac{1}{S} p_n - \frac{1}{S} \{x_{1,n}^2 + x_{2,n}^2 + (2x_{1,n}x_{2,n} + x_{1,n}x_{3,n} + x_{2,n}x_{4,n}) \\
&\quad (1-S-\Sigma x_{i,n}^2) V^{-1}\} \\
q_{n+1} &= \frac{1}{S} q_n - \frac{1}{S} \{x_{1,n}^2 + x_{2,n}^2 + (2x_{1,n}x_{3,n} + x_{1,n}x_{2,n} + x_{3,n}x_{4,n}) \\
&\quad (1-S-\Sigma x_{i,n}^2) V^{-1}\} ,
\end{aligned} \quad (3.104)$$

where V is defined in (3.101). Consider the difference

$$q_{n+1} - p_{n+1} = \frac{1}{S} (q_n - p_n) (x_{1,n} + x_{4,n}) [1 - (1 - S - \Sigma x_{i,n}^2) V^{-1}]. \quad (3.105)$$

When $q_0 = p_0$, it is seen from (3.105) that $q = p$ in every generation. If $q_0 \neq p_0$, the difference $(q_n - p_n)$ will be increased by selection. Note here that, from the defining inequality for this stage, the term $[1 - (1 - S - \Sigma x_{i,n}^2) V^{-1}]$ appearing in (3.105) is greater than zero. For the symmetrical case $x_1 = x_4$ and $x_2 = x_3$, equations (3.105) reduce to

$$p_{n+1} = q_{n+1} = \frac{1}{2}.$$

The gametic disequilibrium in the $(n+1)$ th generation is given by

$$D_{n+1} = x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1}.$$

Using equations (3.103) for the symmetrical case, we find

$$D_{n+1} = \frac{1}{S^2} \left(\frac{1}{2} - r\right)^2 D_n - \frac{1}{2} \left(\frac{1}{2} - r\right) \left(\frac{1}{2} - r - S\right) (x_{1,n} - x_{2,n}). \quad (3.106)$$

Equation (3.106) shows that when $r = \frac{1}{2}$ the population will be in gametic equilibrium irrespective of selection.

The change in the gametic frequencies is given by

$$\Delta x_{i,n} = x_{i,n+1} - x_{i,n} \quad i = 1, 2, 3, 4.$$

Using equations (3.103) for the symmetrical case, we find

$$\Delta x_{i,n} = \frac{1}{S} \left(\frac{1}{2} - r - S\right) \left(x_{i,n} - \frac{1}{4}\right) \quad i = 1, 2, 3, 4. \quad (3.107)$$

Equilibrium exists when $\Delta x_i = 0$ for all i . It is seen from (3.107) that equilibrium exists for $x_1 = x_2 = x_3 = x_4 = \frac{1}{4}$. The change in gametic frequencies may be continuized by writing

$$\Delta x_{i,n} = \frac{1}{S} \left(\frac{1}{2} - r-S \right) \left(x_{i,n} - \frac{1}{4} \right) \doteq \frac{dx_i}{dn} . \quad (3.108)$$

Equation (3.108) yields the following solution

$$\ln \left| x_{i,n} - \frac{1}{4} \right| = \ln \left| x_{i,0} - \frac{1}{4} \right| + \frac{1}{S} \left(\frac{1}{2} - r-S \right) n , \quad (3.109)$$

for $i=1,2,3,4$.

$$3. \quad (1-S) \geq 1 - 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n}).$$

This stage is the same as the second stage discussed above for the overdominance model with interaction. Thus equations (3.73) to (3.90) apply to this stage.

Two loci with unequal effects

In this section the two loci A and B are assumed to have unequal effects with the B gene more advantageous than the A gene.

No dominance of both loci. The mean phenotypic values for the genotypes are given in Table 20. The numerical values in the body of the table determine only the order of the genotypes. The mathematics that follows does not depend on the particular values used.

Table 20. Mean phenotypic values of the genotypes.

	AA	Aa	aa
BB	13	10.5	8
Bb	9	6.5	4
bb	5	2.5	0

The order of the genotypes is

$$aabb < Aabb < aaBb < AAbb < AaBb < aaBB < AABb < AaBB < AABB .$$

There are nine possible stages of progress.

$$1. \quad (1-S) < x_{4,n}^2.$$

This stage is the same as the first stage discussed above under the model of no dominance with equal effects at the two loci. Thus equations (3.1) to (3.16) apply to this stage.

$$2. \quad x_{4,n}^2 \leq (1-S) < x_{4,n}^2 + 2x_{2,n}x_{4,n} .$$

We find

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} \{x_{1,n}^2 + x_{1,n}x_{3,n} + x_{1,n}x_{2,n} + rx_{2,n}x_{3,n} + (1-r)x_{1,n}x_{4,n}\} \\ &= \frac{1}{S} (x_{1,n} - rD_n) \end{aligned}$$

$$\begin{aligned} x_{2,n+1} &= \frac{1}{S} \{x_{1,n}x_{2,n} + (1-r)x_{2,n}x_{3,n} + rx_{1,n}x_{4,n} + x_{2,n}^2 \\ &\quad + \frac{1}{2} [2x_{2,n}x_{4,n} - (1-S-x_{4,n}^2)]\} \\ &= \frac{1}{S} [x_{2,n} + rD_n - \frac{1}{2} (1-S-x_{4,n}^2)] \end{aligned} \quad (4.1)$$

$$\begin{aligned}
 x_{3,n+1} &= \frac{1}{S} \{x_{1,n}x_{3,n} + (1-r)x_{2,n}x_{3,n} + rx_{1,n}x_{4,n} + x_{3,n}^2 + x_{3,n}x_{4,n}\} \\
 &= \frac{1}{S} (x_{3,n} + rD_n)
 \end{aligned}$$

$$\begin{aligned}
 x_{4,n+1} &= \frac{1}{S} \{ (1-r)x_{1,n}x_{4,n} + rx_{2,n}x_{3,n} + x_{3,n}x_{4,n} \\
 &\quad + \frac{1}{2} [2x_{2,n}x_{4,n} - (1-S-x_{4,n}^2)] \} \\
 &= \frac{1}{S} [x_{4,n}(1-x_{4,n}) - rD_n - \frac{1}{2} (1-S-x_{4,n}^2)] .
 \end{aligned}$$

The gametic disequilibrium in the (n+1)th generation is given by

$$D_{n+1} = x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1} .$$

Using (4.1), we find

$$D_{n+1} = \frac{1-rS}{S^2} D_n + \frac{1}{2S^2} (x_{1,n} + x_{3,n})(1-S-x_{4,n}^2) . \quad (4.2)$$

So, as long as the inequality on S holds,

$$\begin{aligned}
 D_n &= \left(\frac{1-rS}{S^2}\right)^n D_0 + \frac{1}{2(1-rS)} \sum_{i=0}^{n-1} \left(\frac{1-rS}{S^2}\right)^{n-1-i} \\
 &\quad (x_{1,i} + x_{3,i})(1-S-x_{4,i}^2) \quad (4.3)
 \end{aligned}$$

From the defining inequality for the second stage, the term $(1-S-x_{4,i}^2)$ is greater than zero and the second term on the right-hand side of (4.3) is greater than zero. A comparison between (4.3) and (3.21) shows that the gametic disequilibrium built up by selection is higher during the second stage when the two loci have unequal effects than when their effects are

equal. Equation (4.3) shows that selection may build up positive gametic disequilibrium even when the initial population is in gametic equilibrium, i.e. $D_0 = 0$.

The new frequencies of the A and B genes are given by

$$p_{n+1} = x_{1,n+1} + x_{2,n+1}$$

$$q_{n+1} = x_{1,n+1} + x_{3,n+1}$$

respectively. Using (4.1), we find

$$\begin{aligned} p_{n+1} &= \frac{1}{S} p_n - \frac{1}{2S} (1-S-x_{4,n}^2) \\ q_{n+1} &= \frac{1}{S} q_n \end{aligned} \quad (4.4)$$

We now consider the difference

$$q_{n+1} - p_{n+1} = \frac{1}{S} (q_n - p_n) + \frac{1}{2S} (1-S-x_{4,n}^2). \quad (4.5)$$

So, as long as the inequality on S holds,

$$q_n - p_n = \frac{1}{S^n} (q_0 - p_0) + \frac{1}{2} \sum_{i=0}^{n-1} S^{-n+i} (1-S-x_{4,i}^2). \quad (4.6)$$

Equation (4.6) shows that, even when $q_0 = p_0$, the frequency of the B gene will be increased by selection over the A gene because $(1-S) > x_{4,i}^2$ during this stage.

$$3. \quad x_{4,n}^2 + 2x_{2,n}x_{4,n} \leq (1-S) < x_{4,n}^2 + 2x_{2,n}x_{4,n} + 2x_{3,n}x_{4,n}.$$

We find

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} \{x_{1,n}^2 + x_{1,n}x_{2,n} + x_{1,n}x_{3,n} + rx_{2,n}x_{3,n} + (1-r)x_{1,n}x_{4,n}\} \\ &= \frac{1}{S} (x_{1,n} - rD_n) \end{aligned}$$

$$\begin{aligned}
x_{2,n+1} &= \frac{1}{S} \{x_{1,n}x_{2,n} + (1-r)x_{2,n}x_{3,n} + rx_{1,n}x_{4,n} + x_{2,n}^2\} \\
&= \frac{1}{S} [x_{2,n}(1-x_{4,n}) + rD_n]
\end{aligned} \tag{4.7}$$

$$\begin{aligned}
x_{3,n+1} &= \frac{1}{S} \{x_{1,n}x_{3,n} + (1-r)x_{2,n}x_{3,n} + rx_{1,n}x_{4,n} + x_{3,n}^2 \\
&\quad + \frac{1}{2} [2x_{3,n}x_{4,n} - (1-S-x_{4,n}^2 - 2x_{2,n}x_{4,n})]\} \\
&= \frac{1}{S} [x_{3,n} + rD_n - \frac{1}{2} (1-S-x_{4,n}^2 - 2x_{2,n}x_{4,n})]
\end{aligned}$$

$$\begin{aligned}
x_{4,n+1} &= \frac{1}{S} \{(1-r)x_{1,n}x_{4,n} + rx_{2,n}x_{3,n} + \frac{1}{2} [2x_{3,n}x_{4,n} \\
&\quad - (1-S-x_{4,n}^2 - 2x_{2,n}x_{4,n})]\} \\
&= \frac{1}{S} [x_{4,n} (1 - \frac{1}{2} x_{4,n}) - \frac{1}{2} (1-S) - rD_n] .
\end{aligned}$$

The gametic disequilibrium in the (n+1)th generation is given by

$$D_{n+1} = x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1} .$$

Using (4.7), we find

$$\begin{aligned}
D_{n+1} &= \frac{1}{S^2} (1-rS - \frac{1}{2} x_{4,n}) D_n \\
&\quad - \frac{1}{2S^2} (1-S) [x_{1,n} - x_{2,n}(1-x_{4,n})] \\
&\quad - x_{2,n}x_{4,n} [(1-x_{4,n})(x_{2,n} + \frac{1}{2} x_{4,n}) - \frac{1}{2} x_{3,n}] .
\end{aligned} \tag{4.8}$$

The new frequencies of the A and B genes are given by

$$\begin{aligned}
p_{n+1} &= x_{1,n+1} + x_{2,n+1} \\
q_{n+1} &= x_{1,n+1} + x_{3,n+1} ,
\end{aligned}$$

respectively. Using (4.7), we find

$$\begin{aligned} p_{n+1} &= \frac{1}{S} p_n - \frac{1}{S} x_{2,n} x_{4,n} \\ q_{n+1} &= \frac{1}{S} q_n - \frac{1}{2S} (1-S-x_{4,n}^2 - 2x_{2,n} x_{4,n}) . \end{aligned} \quad (4.9)$$

From (4.9) we have

$$q_{n+1} - p_{n+1} = \frac{1}{S} (q_n - p_n) + \frac{1}{2S} [x_{4,n}^2 + 4x_{2,n} x_{4,n} - (1-S)] . \quad (4.10)$$

From the defining inequality for the third stage it is seen from (4.10) that, when $x_{2,n} = x_{3,n}$, the second term on the right-hand side of (4.10) is positive and selection will continue to increase the difference between the frequencies of the two genes. A comparison between (4.5) and (4.10) shows that the rate of increase in the difference between the frequencies of the two genes is higher in the second stage than in the third stage.

$$4. \quad x_{4,n}^2 + 2x_{2,n} x_{4,n} + 2x_{3,n} x_{4,n} \leq (1-S) < x_{4,n}^2 + 2x_{2,n} x_{4,n} + 2x_{3,n} x_{4,n} + x_{2,n}^2 .$$

We find

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} \{x_{1,n}^2 + x_{1,n} x_{3,n} + x_{1,n} x_{2,n} + (1-r)x_{1,n} x_{4,n} + rx_{2,n} x_{3,n}\} \\ &= \frac{1}{S} (x_{1,n} - rD_n) \\ x_{2,n+1} &= \frac{1}{S} \{x_{1,n} x_{2,n} + (1-r)x_{2,n} x_{3,n} + rx_{1,n} x_{4,n} + x_{2,n}^2 \\ &\quad - (1-S-x_{4,n}^2 - 2x_{2,n} x_{4,n} - 2x_{3,n} x_{4,n})\} \\ &= \frac{1}{S} [x_{2,n} + rD_n - (1-S) + x_{4,n}(1-x_{1,n} + x_{3,n})] \end{aligned} \quad (4.11)$$

$$\begin{aligned}
 x_{3,n+1} &= \frac{1}{S} \{x_{1,n}x_{3,n} + x_{3,n}^2 + rx_{1,n}x_{4,n} + (1-r)x_{2,n}x_{3,n}\} \\
 &= \frac{1}{S} [x_{3,n}(1-x_{4,n}) + rD_n]
 \end{aligned}$$

$$\begin{aligned}
 x_{4,n+1} &= \frac{1}{S} \{(1-r)x_{1,n}x_{4,n} + rx_{2,n}x_{3,n}\} \\
 &= \frac{1}{S} [x_{1,n}x_{4,n} - rD_n] .
 \end{aligned}$$

The gametic disequilibrium in the $(n+1)$ th generation is given by

$$D_{n+1} = x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1} .$$

Using (4.11), we find

$$\begin{aligned}
 D_{n+1} &= \frac{1}{S^2} (1-rS-x_{4,n})D_n - x_{1,n}x_{4,n}(x_{2,n} + x_{3,n}x_{4,n}) \\
 &\quad + x_{3,n}(1-x_{4,n})[1-S-x_{4,n}(1+x_{3,n})] . \quad (4.12)
 \end{aligned}$$

Equation (4.12) shows that selection will continue to build up gametic disequilibrium even when $D_0 = 0$.

The new frequencies of the A and B genes are given by

$$\begin{aligned}
 p_{n+1} &= x_{1,n+1} + x_{2,n+1} \\
 q_{n+1} &= x_{1,n+1} + x_{3,n+1} .
 \end{aligned}$$

Using (4.11), we find

$$\begin{aligned}
 p_{n+1} &= \frac{1}{S} p_n - \frac{1}{S} [1-S-x_{4,n}(1-x_{1,n} + x_{3,n})] \\
 q_{n+1} &= \frac{1}{S} q_n - \frac{1}{S} x_{3,n}x_{4,n} . \quad (4.13)
 \end{aligned}$$

From (4.13) we have

$$q_{n+1} - p_{n+1} = \frac{1}{S} (q_n - p_n) + \frac{1}{S} [1 - S - x_{4,n}(1 - x_{1,n} + x_{3,n}) - x_{3,n}x_{4,n}]. \quad (4.14)$$

So, as long as the inequality on S holds,

$$q_n - p_n = \frac{1}{S^n} (q_0 - p_0) + \sum_{i=0}^{n-1} \frac{1}{S^{n-i}} [1 - S - x_{4,i}(1 - x_{1,i} + x_{3,i}) - x_{3,i}x_{4,i}]. \quad (4.15)$$

Equation (4.14) shows the difference $(q_n - p_n)$ will be increased by selection during the fourth stage even when $q_0 = p_0$. A comparison between (4.10) and (4.14) shows that the rate of increase in the difference between the frequencies of the two genes is higher in the fourth than in the third stage if

$$1 - S > x_{4,n}^2 + 2x_{2,n}x_{4,n} + 2x_{3,n}x_{4,n},$$

a condition which is satisfied from the defining inequality for the fourth stage.

$$5. \quad x_{4,n}^2 + 2x_{2,n}x_{4,n} + 2x_{3,n}x_{4,n} + x_{2,n}^2 \leq (1 - S) < 1 - (x_{3,n}^2 + 2x_{1,n}x_{2,n} + 2x_{1,n}x_{3,n} + x_{1,n}^2).$$

In this stage all of the AABB, AaBB, AABb, aaBB and some of the AaBb individuals are saved. The array among the AaBb individuals culled in generation n is

$$\frac{2x_{1,n}x_{4,n} \text{ AB/ab} + 2x_{2,n}x_{3,n} \text{ Ab/aB}}{2x_{1,n}x_{4,n} + 2x_{2,n}x_{3,n}}.$$

The array among the saved individuals in the same generation is

$$\begin{aligned}
& \frac{1}{S} \{x_{1,n}^2 \text{ AABB} + 2x_{1,n}x_{2,n} \text{ AABb} + 2x_{1,n}x_{3,n} \text{ AaBB} + x_{3,n}^2 \text{ aaBB} \\
& + 2x_{1,n}x_{4,n} \left[1 - \frac{1}{2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n})} (1-S-x_{4,n}^2 - 2x_{2,n}x_{4,n} \right. \\
& \left. - 2x_{3,n}x_{4,n} - x_{2,n}^2) \right] \text{ AB/ab} + 2x_{2,n}x_{3,n} \left[1 - \frac{1}{2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n})} \right. \\
& \left. (1-S-x_{4,n}^2 - 2x_{2,n}x_{4,n} - 2x_{3,n}x_{4,n} - x_{2,n}^2) \right] \text{ Ab/aB} \} .
\end{aligned}$$

Thus we have

$$\begin{aligned}
x_{1,n+1} &= \frac{1}{S} \{x_{1,n}^2 + x_{1,n}x_{2,n} + x_{1,n}x_{3,n} + (1-r)x_{1,n}x_{4,n}(1-W_n) \\
& \quad + rx_{2,n}x_{3,n}(1-W_n)\} \\
&= \frac{1}{S} [x_{1,n} - (1-W_n) rD_n - x_{1,n}x_{4,n}W_n] \\
x_{2,n+1} &= \frac{1}{S} \{x_{1,n}x_{2,n} + rx_{1,n}x_{4,n}(1-W_n) + (1-r)x_{2,n}x_{3,n}(1-W_n)\} \\
&= \frac{1}{S} [x_{2,n}(x_{1,n} + x_{3,n}) + rD_n(1-W_n) - x_{2,n}x_{3,n}W_n] \\
x_{3,n+1} &= \frac{1}{S} \{x_{1,n}x_{3,n} + x_{3,n}^2 + rx_{1,n}x_{4,n}(1-W_n) + (1-r)x_{2,n}x_{3,n}(1-W_n)\} \\
&= \frac{1}{S} [x_{3,n}(1-x_{4,n}) + rD_n(1-W_n) - x_{2,n}x_{3,n}W_n] \\
x_{4,n+1} &= \frac{1}{S} \{(1-r)x_{1,n}x_{4,n}(1-W_n) + rx_{2,n}x_{3,n}(1-W_n)\} \\
&= \frac{1}{S} (x_{1,n}x_{4,n} - rD_n)(1-W_n) ,
\end{aligned} \tag{4.16}$$

where

$$W_n = [2x_{1,n}x_{4,n} + 2x_{2,n}x_{3,n}]^{-1} (1 - s - x_{4,n}^2 - 2x_{2,n}x_{4,n} - 2x_{3,n}x_{4,n} - x_{2,n}^2) . \quad (4.17)$$

The new frequencies of the A and B genes are given by

$$p_{n+1} = x_{1,n+1} + x_{2,n+1}$$

$$q_{n+1} = x_{1,n+1} + x_{3,n+1} ,$$

respectively. Using (4.16) we find

$$p_{n+1} = \frac{1}{S} p_n - \frac{1}{2S} (1 - s - x_{4,n}^2 - 2x_{3,n}x_{4,n} + x_{2,n}^2)$$

$$q_{n+1} = \frac{1}{S} q_n - \frac{1}{2S} (1 - s - x_{4,n}^2 - 2x_{2,n}x_{4,n} - x_{2,n}^2) . \quad (4.18)$$

From (4.18) we have

$$q_{n+1} - p_{n+1} = \frac{1}{S} (q_n - p_n) + \frac{1}{S} [x_{2,n}^2 - x_{4,n}(q_n - p_n)] . \quad (4.19)$$

If $q_n = p_n$, it is seen from (4.19) that the difference between the two gene frequencies will be increased by selection.

$$6. \quad 1 - (x_{3,n}^2 + 2x_{1,n}x_{2,n} + 2x_{1,n}x_{3,n} + x_{1,n}^2) \leq (1-s) < 1 - (x_{1,n}^2 + 2x_{1,n}x_{2,n} + 2x_{1,n}x_{3,n}) .$$

In this stage we have $x_1 + x_2 + x_3 = 1$ and $x_4 = 0$. We find

$$\begin{aligned}
x_{1,n+1} &= \frac{1}{S} (x_{1,n}^2 + x_{1,n}x_{2,n} + x_{1,n}x_{3,n}) \\
&= \frac{1}{S} x_{1,n} \\
x_{2,n+1} &= \frac{1}{S} x_{1,n}x_{2,n} \\
x_{3,n+1} &= \frac{1}{S} \{x_{1,n}x_{3,n} + x_{3,n}^2 - [1-S(1-x_{3,n}^2 - 2x_{1,n}x_{2,n} \\
&\quad - 2x_{1,n}x_{3,n} - x_{1,n}^2)]\} \\
&= \frac{1}{S} (S - x_{1,n} - x_{1,n}x_{2,n}) \\
x_{4,n+1} &= 0 .
\end{aligned} \tag{4.20}$$

The new frequencies of the A and B genes are given by

$$\begin{aligned}
p_{n+1} &= x_{1,n+1} + x_{2,n+1} \\
q_{n+1} &= x_{1,n+1} + x_{3,n+1} ,
\end{aligned}$$

respectively. Using (4.20), we find

$$\begin{aligned}
p_{n+1} &= \frac{1}{S} p_n - \frac{1}{S} x_{2,n}(1-x_{1,n}) \\
q_{n+1} &= 1 - \frac{1}{S} x_{1,n}x_{2,n} .
\end{aligned} \tag{4.21}$$

It is from (4.21) that gene frequencies will continue to change and that the difference $(q_{n+1} - p_{n+1})$ will be zero when $x_2 = 0$ and $x_1 = 1$.

$$7. \quad 1 - (x_{1,n}^2 + 2x_{1,n}x_{3,n} + 2x_{1,n}x_{2,n}) \leq (1-S) < 1 - (x_{1,n}^2 + 2x_{1,n}x_{3,n}).$$

In this stage we have $x_1 + x_2 + x_3 = 1$ and $x_4 = 0$. We find

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} \{x_{1,n}^2 + x_{1,n}x_{3,n} + x_{1,n}x_{2,n} - \frac{1}{2} [1-S-(1-x_{1,n}^2 \\ &\quad - 2x_{1,n}x_{2,n} - 2x_{1,n}x_{3,n})]\} \\ &= \frac{1}{2S} (S + x_{1,n}^2) \\ x_{2,n+1} &= \frac{1}{S} \{x_{1,n}x_{2,n} - \frac{1}{2} [1-S-(1-x_{1,n}^2 - 2x_{1,n}x_{2,n} - 2x_{1,n}x_{3,n})]\} \\ &= \frac{1}{2S} (S - x_{1,n}^2 - 2x_{1,n}x_{3,n}) \end{aligned} \quad (4.22)$$

$$x_{3,n+1} = \frac{1}{S} x_{1,n}x_{3,n}$$

$$x_{4,n+1} = 0.$$

The new frequencies of the A and B genes are

$$\begin{aligned} p_{n+1} &= x_{1,n+1} + x_{2,n+1} = 1 - \frac{1}{S} x_{1,n}x_{2,n} \\ q_{n+1} &= x_{1,n+1} + x_{3,n+1} = \frac{1}{2S} [S + x_{1,n}(2-x_{1,n} - x_{2,n})]. \end{aligned} \quad (4.23)$$

It is seen from (4.23) that the difference $(q_{n+1} - p_{n+1})$ will be zero when $x_1 = S$ and $x_2 = 0$.

$$8. \quad 1 - (x_{1,n}^2 + 2x_{1,n}x_{3,n}) \leq (1-S) < 1 - x_{1,n}^2.$$

In this stage we have $x_1 + x_3 = 1$ and $x_2 = x_4 = 0$. We find

$$\begin{aligned}
 x_{1,n+1} &= \frac{1}{S} \{x_{1,n}^2 + x_{1,n}x_{3,n} - \frac{1}{2} [1-S-(1-x_{1,n}^2 - 2x_{1,n}x_{3,n})]\} \\
 &= \frac{1}{2S} (S + x_{1,n}^2)
 \end{aligned}$$

$$x_{2,n+1} = 0$$

(4.24)

$$\begin{aligned}
 x_{3,n+1} &= \frac{1}{S} \{x_{1,n}x_{3,n} - \frac{1}{2} [1-S-(1-x_{1,n}^2 - 2x_{1,n}x_{3,n})]\} \\
 &= \frac{1}{S} (S-x_{1,n}^2)
 \end{aligned}$$

$$x_{4,n+1} = 0.$$

It is seen from (4.24) that as x_1 approaches \sqrt{S} the frequency x_3 will approach zero and $q_{n+1} = p_{n+1} = 1$.

$$9. \quad (1-S) \geq 1 - x_{1,n}^2.$$

After one generation of selection the population will consist solely of AABB individuals and we have at equilibrium $x_1 = 1$, $x_2 = x_3 = x_4 = 0$ and $p = q = 1$.

VARIABILITY OF EXPRESSION OF GENETIC FACTORS

One-Locus Case

The model that will be assumed in this section is as follows. We consider the case of a single locus with two alleles A, a and suppose that the expressions of the genotypes are variable so that the value on the scale of merit (t) of an individual with genotype $A_i A_j$ is normally distributed around some fixed value μ_{ij} with variance σ_{ij}^2 , independently from individual to individual. We assume that the population is very large so that all fluctuations due to sampling from a finite population may be ignored. We assumed that in any generation, the individuals which are in the top fraction S of the whole population of phenotypic values are saved to produce the next generation, with S being constant. Under this type of selection, the fractions selected from each phenotypic distribution will vary from generation to generation, and there may be further selection among those not initially culled. This type of model raises several interesting questions, some of which we enumerate briefly:

- (i) What is the progress of the population with given S , μ_{ij} , and σ_{ij}^2 ?
- (ii) What are the relationships between the effects of the means μ_{ij} and the variances σ_{ij}^2 on the equilibrium that is attained and the progress of the population toward its equilibrium?

Haldane (1931) considered the problem of truncation selection in which a character measured by x was assumed to be normally distributed with the amount of variability being the same in each group. The truncation

point x_0 , say, is assumed to be known. It can be seen from his analysis that the efficiency of selection increases very rapidly with quantity z up to a point after which the efficiency declines. This quantity z corresponds to the ratio $(1-S)/S$ which is the ratio of the proportion culled to the proportion saved.

Robertson (1956) and Curnow (1964) examined a different type of selection. In their experiment, individuals with values of a certain character near the population mean are selected. In other words, selection applied was within a symmetric interval from the population mean with the interval points being known. A consequence of this selection scheme is that the proportion culled varies with the gene frequency. A more interesting case would be that in which there is selection for an intermediate with a fixed culling proportion. This could be represented by using a scale of merit equal to the negative of $|x-\mu|$ or $(x-\mu)^2$, where μ is the population mean. With additive gene effects on x this would amount to overdominance on the scale of merit.

With one locus and two alleles A_1 and A_0 there are three possible genotypes designated A_1A_1 , A_1A_0 and A_0A_0 . The genotypic array in generation n before selection is

$$p_n^2 A_1A_1 + 2p_n(1-p_n) A_1A_0 + (1-p_n)^2 A_0A_0 ,$$

where p_n is the frequency of the A_1 gene in generation n . We denote by μ_{ij} and σ_{ij}^2 the mean phenotypic value and the phenotypic variance associated with the genotype A_iA_j ($i = 0,1$ and $j = 0,1$), respectively.

Under the assumption of normality, the probability that an individual of genotype $A_i A_j$ survives selection is given by

$$\text{Prob}(t > x_0) = \frac{1}{\sigma_{ij}\sqrt{2\pi}} \int_{x_0}^{\infty} \exp \left[-\frac{(t-\mu_{ij})^2}{2\sigma_{ij}^2} \right] dt ,$$

where t is the variate and x_0 is the truncation point. The proportion of individuals saved in generation n is

$$\begin{aligned} S = & p_n^2 \frac{1}{\sigma_{11}\sqrt{2\pi}} \int_{x_n}^{\infty} \exp \left[-\frac{(t-\mu_{11})^2}{2\sigma_{11}^2} \right] dt \\ & + 2p_n(1-p_n) \frac{1}{\sigma_{10}\sqrt{2\pi}} \int_{x_n}^{\infty} \exp \left[-\frac{(t-\mu_{10})^2}{2\sigma_{10}^2} \right] dt \\ & + (1-p_n)^2 \frac{1}{\sigma_{00}\sqrt{2\pi}} \int_{x_n}^{\infty} \exp \left[-\frac{(t-\mu_{00})^2}{2\sigma_{00}^2} \right] dt . \end{aligned} \quad (5.1)$$

We now take up the case in which $\sigma_{ij} = \sigma$ is the same for all genotypes.

Let

$$N_{ij}(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(t-\mu_{ij})^2}{2\sigma^2} \right] \quad (5.2)$$

$$\begin{aligned} M_{ij}(x_n) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{x_n}^{\infty} \exp \left[-\frac{(t-\mu_{ij})^2}{2\sigma^2} \right] dt \\ &= \int_{x_n}^{\infty} N_{ij}(t) dt . \end{aligned} \quad (5.3)$$

Equation (5.1) may be written in the form

$$S = p_n^2 M_{11}(x_n) + 2p_n(1-p_n) M_{10}(x_n) + (1-p_n)^2 M_{00}(x_n) . \quad (5.4)$$

Then, using a Taylor's series expansion around μ_{10} , we find

$$\begin{aligned} M_{11}(x_n) &= M_{10}(x_n) + (\mu_{11} - \mu_{10}) M'_{10}(x_n) + \frac{1}{2} (\mu_{11} - \mu_{10})^2 M''_{10}(x_n) + \dots \\ M_{00}(x_n) &= M_{10}(x_n) + (\mu_{00} - \mu_{10}) M'_{10}(x_n) + \frac{1}{2} (\mu_{00} - \mu_{10})^2 M''_{10}(x_n) + \dots \end{aligned} \quad (5.5)$$

where the primes denote the degree of derivative. If the differences between the μ_{ij} 's are small such that orders higher than the first can be ignored, equations (5.5) may be approximated by

$$\begin{aligned} M_{11}(x_n) &= M_{10}(x_n) + (\mu_{11} - \mu_{10}) M'_{10}(x_n) \\ M_{00}(x_n) &= M_{10}(x_n) + (\mu_{00} - \mu_{10}) M'_{10}(x_n) . \end{aligned} \quad (5.6)$$

Differentiating $M_{10}(x_n)$ given by (5.3) within the integral we find

$$\begin{aligned} M'_{10}(x_n) &= \frac{d}{d\mu_{10}} M_{10}(x_n) \\ &= \frac{d}{d\mu_{10}} \frac{1}{\sigma\sqrt{2\pi}} \int_{x_n}^{\infty} \exp \left[-\frac{(t-\mu_{10})^2}{2\sigma^2} \right] dt \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{x_n}^{\infty} \frac{(t-\mu_{10})}{\sigma^2} \exp \left[-\frac{(t-\mu_{10})^2}{2\sigma^2} \right] dt \\ &= \frac{1}{\sigma\sqrt{2\pi}} \left\{ -\exp \left[-\frac{(t-\mu_{10})^2}{2\sigma^2} \right] \right\}_{x_n}^{\infty} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x_n - \mu_{10})^2}{2\sigma^2} \right] \\
&= N_{10}(x_n) .
\end{aligned} \tag{5.7}$$

Similarly,

$$\begin{aligned}
M''_{10}(x_n) &= \frac{d}{d\mu_{10}} M'_{10}(x_n) \\
&= \frac{1}{\sigma\sqrt{2\pi}} \frac{(x_n - \mu_{10})}{\sigma^2} \exp \left[-\frac{(x_n - \mu_{10})^2}{2\sigma^2} \right] \\
&= \frac{(x_n - \mu_{10})}{\sigma^2} N_{10}(x_n) .
\end{aligned} \tag{5.8}$$

We also have, approximately,

$$\begin{aligned}
\frac{d}{dn} M_{10}(x_n) &= \frac{dx_n}{dn} \frac{d}{dx_n} M_{10}(x_n) \\
&= -\frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x_n - \mu_{10})^2}{2\sigma^2} \right] \frac{dx_n}{dn} \\
&= -N_{10}(x_n) \frac{dx_n}{dn}
\end{aligned} \tag{5.9}$$

$$\begin{aligned}
\frac{d}{dn} N_{10}(x_n) &= \frac{dx_n}{dn} \frac{d}{dx_n} N_{10}(x_n) \\
&= -\frac{1}{\sigma\sqrt{2\pi}} \frac{(x_n - \mu_{10})}{\sigma^2} \exp \left[-\frac{(x_n - \mu_{10})^2}{2\sigma^2} \right] \frac{dx_n}{dn} \\
&= -\frac{(x_n - \mu_{10})}{\sigma^2} N_{10}(x_n) \frac{dx_n}{dn} .
\end{aligned} \tag{5.10}$$

Replacing $M_{10}^1(x_n)$ in (5.6) by its value from (5.7), we find

$$\begin{aligned} M_{11}(x_n) &= M_{10}(x_n) + (\mu_{11} - \mu_{10}) N_{10}(x_n) \\ M_{00}(x_n) &= M_{10}(x_n) + (\mu_{00} - \mu_{10}) N_{10}(x_n) . \end{aligned} \quad (5.11)$$

Using (5.11), equation (5.4) may be written in the approximate form

$$\begin{aligned} S &= p_n^2 [M_{10}(x_n) + (\mu_{11} - \mu_{10}) N_{10}(x_n)] + 2p_n(1-p_n)M_{10}(x_n) \\ &\quad + (1-p_n)^2 [M_{10}(x_n) + (\mu_{00} - \mu_{10}) N_{10}(x_n)] \\ &= M_{10}(x_n) + [(\mu_{11} - \mu_{10})^2 p_n^2 + (\mu_{00} - \mu_{10})(1-p_n)^2] N_{10}(x_n) . \end{aligned} \quad (5.12)$$

Equations (5.4) and (5.12) may be solved as a quadratic in p_n to give p_n in terms of x_n . From (5.4) we have, using the fact that $p_n \geq 0$,

$$p_n = \frac{M_{00} - M_{10} + \sqrt{(M_{10}^2 - M_{11}M_{00}) + S(M_{11} - 2M_{10} + M_{00})}}{M_{11} - 2M_{10} + M_{00}} , \quad (5.13)$$

where the M_{ij} are functions of x_n . With no overdominance, the signs of the first and second terms under the square root are negative and positive, respectively, and the two terms interchange signs when overdominance is present. For both cases, in order for the solution (5.13) to exist we must have

$$S \geq \frac{M_{10}^2(x_n) - M_{11}(x_n) M_{00}(x_n)}{M_{11}(x_n) - 2M_{10}(x_n) + M_{00}(x_n)} ,$$

a condition which is satisfied for all $S(0 < S < 1)$. From (5.12) we have

$$p_n = \frac{\mu_{00} - \mu_{10} + \sqrt{(\mu_{11} - \mu_{10})(\mu_{10} - \mu_{00}) + (\mu_{11} - 2\mu_{10} + \mu_{00})(S - M_{10})/N_{10}(x_n)}}{\mu_{11} - 2\mu_{10} + \mu_{00}} \quad (5.14)$$

where both M_{10} and N_{10} are functions of x_n . With no dominance, i.e.

$\mu_{11} - \mu_{10} = \mu_{10} - \mu_{00}$, it is seen from (5.12) that the solution is

$$p_n = \frac{S - M_{10}(x_n)}{2(\mu_{10} - \mu_{00})N_{10}(x_n)} + \frac{1}{2}.$$

The gene frequency after one generation of selection is given by

$$p_{n+1} = \frac{1}{S} [p_n^2 M_{11}(x_n) + p_n(1-p_n) M_{10}(x_n)].$$

Replacing $M_{11}(x_n)$ by its value from (5.11) we find, approximately,

$$\begin{aligned} p_{n+1} &= \frac{1}{S} [p_n^2 M_{10}(x_n) + (\mu_{11} - \mu_{10}) N_{10}(x_n) + p_n(1-p_n) M_{10}(x_n)] \\ &= \frac{1}{S} [(\mu_{11} - \mu_{10}) N_{10}(x_n) p_n^2 + p_n M_{10}(x_n)]. \end{aligned} \quad (5.15)$$

The new gene frequency given by (5.15) may also be written in the form

$$\begin{aligned} p_{n+1} &= p_n + \frac{1}{S} p_n(1-p_n) N_{10}(x_n) [p_n(\mu_{11} - \mu_{10}) + (1-p_n)(\mu_{10} - \mu_{00})] \\ &= p_n + \Delta p_n, \end{aligned} \quad (5.16)$$

where

$$\begin{aligned} \Delta p_n &= \frac{1}{S} p_n(1-p_n) N_{10}(x_n) [p_n(\mu_{11} - \mu_{10}) + (1-p_n)(\mu_{10} - \mu_{00})] \\ &= \frac{1}{S} p_n(1-p_n) N_{10}(x_n) [\text{effect of gene substitution}]. \end{aligned} \quad (5.17)$$

For the case of a continuous change of gene frequency with time, we may write $\Delta p_n = (dp/dn)$, and we have

$$\frac{dp}{dn} = \frac{1}{S} p_n (1-p_n) N_{10}(x_n) [p_n (\mu_{11} - \mu_{10}) + (1-p_n) (\mu_{10} - \mu_{00})] . \quad (5.18)$$

The above non-linear differential equation (5.18) may be written in the form

$$dn = \frac{S}{p_n (1-p_n) N_{10}(x_n) [p_n (\mu_{11} - \mu_{10}) + (1-p_n) (\mu_{10} - \mu_{00})]} dp . \quad (5.19)$$

Equation (5.17) or (5.18) expresses the change in gene frequency as a function of S , μ_{ij} , σ , x and p . The non-linear differential equation given by (5.19) is a function of both p and x , each of which is a function of the other. A solution to this differential equation can be obtained if p_n is expressed as a function of x_n or vice versa. An exact solution to (5.19) giving a form of the relationship between p_n and n is very difficult to obtain, and one may have to solve the equation numerically on the computer rather than mathematically. An approximate solution may be obtained as follows. With S being constant every generation, we have

$$\frac{d}{dn} S = 0 ,$$

and the derivative of (5.12) with respect to n yields

$$0 = \frac{d}{dn} M_{10}(x_n) + [p_n^2 (\mu_{11} - \mu_{10}) + (\mu_{00} - \mu_{10}) (1-p_n)^2] \frac{d}{dn} N_{10}(x_n) \\ + N_{10}(x_n) [2(\mu_{11} - \mu_{10}) p_n - 2(\mu_{00} - \mu_{10}) (1-p_n)] \frac{dp}{dn} .$$

Using (5.9) and (5.10), we find

$$\begin{aligned}
 0 &= -N_{10}(x_n) \frac{dx_n}{dn} - \frac{(x_n - \mu_{10})}{\sigma^2} N_{10}(x_n) [p_n^2(\mu_{11} - \mu_{10}) + (1 - p_n)^2(\mu_{00} - \mu_{10})] \frac{dx_n}{dn} \\
 &\quad + N_{10}(x_n) [2(\mu_{11} - \mu_{10})p_n - 2(\mu_{00} - \mu_{10})(1 - p_n)] \frac{dp_n}{dn} \\
 &= -dx_n - \frac{(x_n - \mu_{10})}{\sigma^2} [(\mu_{11} - \mu_{10})p_n^2 + (\mu_{00} - \mu_{10})(1 - p_n)^2] dx_n \\
 &\quad + [2(\mu_{11} - \mu_{10})p_n - 2(\mu_{00} - \mu_{10})(1 - p_n)] dp_n . \tag{5.20}
 \end{aligned}$$

An exact solution to the non-linear differential equation (5.20) is very difficult to obtain mathematically. However, two extreme cases may be considered.

Case 1: σ is very large.

In this case the term $(x_n - \mu_{10})/\sigma^2$ is assumed to be very small. Consequently, equation (5.20) may be written in the approximate form

$$\begin{aligned}
 dx_n &= 2[(\mu_{11} - \mu_{10})p_n - (\mu_{00} - \mu_{10})(1 - p_n)] dp_n \\
 &= 2[(\mu_{11} - 2\mu_{10} + \mu_{00})p_n + (\mu_{10} - \mu_{00})] dp_n \tag{5.21}
 \end{aligned}$$

Equation (5.21) yields the following solution

$$\begin{aligned}
 x_n - x_0 &= (\mu_{11} - 2\mu_{10} + \mu_{00})p_n^2 + (\mu_{10} - \mu_{00})p_n \\
 &\quad - (\mu_{11} - 2\mu_{10} + \mu_{00})p_0^2 - (\mu_{10} - \mu_{00})p_0 . \tag{5.22}
 \end{aligned}$$

For the case when σ is very large, equation (5.22) expresses p_n in terms of x_n or vice versa.

Under the assumption, we may write equation (5.2) approximately in the form

$$N_{10}(x_n) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x_n - \mu_{10})^2}{2\sigma^2} \right] \doteq \frac{1}{\sigma\sqrt{2\pi}} ,$$

and equation (5.19) for the change in gene frequency may be written in the approximate form

$$dn = \frac{S \sigma \sqrt{2\pi}}{p_n(1-p_n) [(\mu_{11}-\mu_{10})p_n + (\mu_{10}-\mu_{00})(1-p_n)]} dp_n . \quad (5.23)$$

Equation (5.23) yields the following solution

$$n = S \sigma \sqrt{2\pi} \left\{ \frac{1}{\mu_{10}-\mu_{00}} \ln \frac{p^2}{(\mu_{10}-\mu_{00}) + (\mu_{11}-3\mu_{10}+2\mu_{00})p - (\mu_{11}-2\mu_{10}+\mu_{00})p^2} \right. \\ \left. - \frac{\mu_{11}-3\mu_{10}+2\mu_{00}}{2(\mu_{11}-\mu_{10})(\mu_{10}-\mu_{00})} \ln \frac{(\mu_{11}-2\mu_{10}+\mu_{00})p + (\mu_{10}-\mu_{00})}{(\mu_{11}-2\mu_{10}+\mu_{00})p - (\mu_{11}-2\mu_{10}+\mu_{00})} \right\} \frac{p_n}{p_0} . \quad (5.24)$$

When there is no dominance, i.e. $\mu_{11} - \mu_{10} = \mu_{10} - \mu_{00}$, equations (5.22) and (5.24) become

$$x_n - x_0 = (\mu_{10}-\mu_{00})(p_n-p_0) \quad (5.25)$$

and

$$n = \frac{S \sigma \sqrt{2\pi}}{(\mu_{10}-\mu_{00})} \left[\ln \frac{1-p_0}{p_0} - \ln \frac{1-p_n}{p_n} \right] , \quad (5.26)$$

respectively. If we let $\mu_{00} = 0$, $\mu_{10} = 1$ and $\mu_{11} = 2$, equations (5.25) and (5.26) reduce to

$$x_n - x_0 = p_n - p_0 \quad (5.27)$$

and

$$n = S \sigma \sqrt{2\pi} [\ln p_n - \ln(1-p_n) - \ln p_0 + \ln(1-p_0)], \quad (5.28)$$

respectively. If we let $\mu_{00} = -u\sigma$, $\mu_{10} = 0$ and $\mu_{11} = u\sigma$, we find

$$x_n - x_0 = u\sigma(p_n - p_0) \quad (5.29)$$

and

$$n = S \sqrt{2\pi} \frac{1}{u} \left[\ln \frac{1-p_0}{p_0} - \ln \frac{1-p_n}{p_n} \right]. \quad (5.30)$$

Equation (5.28) may be written in the form

$$p_n = \frac{1}{1 + \frac{1-p_0}{p_0} \exp \left[- \frac{n}{S \sigma \sqrt{2\pi}} \right]}. \quad (5.31)$$

It is seen from (5.31) that when $n = 0$, we have $p_n = p_0$, and as n increases, p_n approaches the value one. Equation (5.31) is a special case of a logistic form. The curve for the change in gene frequency will be spread out when selection is weak and the degree of variability is large; that is, when both S and σ are large.

For the complete dominance model, i.e. $\mu_{11} = \mu_{10} > \mu_{00}$, equations (5.22) and (5.24) become

$$x_n - x_0 = (\mu_{11} - \mu_{00})p_n - 2(\mu_{11} - \mu_{00})p_n^2 + 2(\mu_{11} - \mu_{00})p_0^2 - (\mu_{11} - \mu_{00})p_0 \quad (5.32)$$

and

$$n = \frac{S \sigma \sqrt{2\pi}}{\mu_{11} - \mu_{00}} \left\{ \frac{1}{1-p} - \ln \frac{1-p}{p} \right\} \frac{p_n}{p_0}, \quad (5.33)$$

respectively. If we let $\mu_{00} = 0$, $\mu_{10} = \mu_{11} = 1$, equations (5.32) and (5.33) reduce to

$$x_n - x_0 = p_n - 2p_n^2 + 2p_0^2 - p_0 \quad (5.34)$$

and

$$n = S \sigma \sqrt{2\pi} \left\{ \frac{1}{1-p} - \ln \frac{1-p}{p} \right\} \frac{p_n}{p_0}, \quad (5.35)$$

respectively. The solution given by (5.35) is similar to that obtained under the one-locus case when there is no variability (equation 1.36), with regard to the form of the function of gene frequency.

If we let $\mu_{00} = 0$ and $\mu_{11} = \mu_{10} = k\sigma$, equations (5.32) and (5.33) become

$$x_n - x_0 = k\sigma [p_n - 2p_n^2 + 2p_0^2 - p_0] \quad (5.36)$$

and

$$n = \frac{1}{k} S \sqrt{2\pi} \left\{ \frac{1}{1-p} - \ln \frac{1-p}{p} \right\} \frac{p_n}{p_0}, \quad (5.37)$$

respectively.

Case 2: σ is very small.

In this case we may write (5.20) in the approximate form

$$\begin{aligned} \frac{(x_n - \mu_{10})^2}{\sigma^2} [(\mu_{11} - \mu_{10})p_n^2 + (\mu_{00} - \mu_{10})(1-p_n)^2] dx_n \\ = 2 [(\mu_{11} - \mu_{10})p_n - (\mu_{00} - \mu_{10})(1-p_n)] dp_n. \end{aligned} \quad (5.38)$$

Equation (5.38) may be written in the form

$$\frac{(x_n - \mu_{10})^2}{\sigma^2} dx_n = \frac{2[(\mu_{11} - 2\mu_{10} + \mu_{00})p_n - (\mu_{00} - \mu_{10})] dp_n}{(\mu_{11} - 2\mu_{10} + \mu_{00})p_n^2 - 2(\mu_{00} - \mu_{10})p_n + (\mu_{00} - \mu_{10})}. \quad (5.39)$$

With no overdominance, i.e. $\mu_{00} < \mu_{10} < \mu_{11}$, equation (5.38) yields the following solution

$$\begin{aligned} \frac{(x_n - \mu_{10})^2}{2\sigma^2} - \frac{(x_0 - \mu_{10})^2}{2\sigma^2} = \{ \ln [(\mu_{11} - 2\mu_{10} + \mu_{00})p^2 - 2(\mu_{00} - \mu_{10})p + (\mu_{00} - \mu_{10})] \\ - (\mu_{10} - \mu_{00}) \ln \frac{(\mu_{11} - 2\mu_{10} + \mu_{00})p + (\mu_{10} - \mu_{00}) - \sqrt{w}}{(\mu_{11} - 2\mu_{10} + \mu_{00})p + (\mu_{10} - \mu_{00}) + \sqrt{w}} \} \frac{p_n}{p_0}, \end{aligned} \quad (5.40)$$

where $w = (\mu_{11} - \mu_{10})(\mu_{10} - \mu_{00})$. When overdominance exists, the solution is given by

$$\begin{aligned} \frac{(x_n - \mu_{10})^2}{2\sigma^2} - \frac{(x_0 - \mu_{10})^2}{2\sigma^2} = \{ \ln [(\mu_{11} - 2\mu_{10} + \mu_{00})p^2 - 2(\mu_{00} - \mu_{10})p + (\mu_{00} - \mu_{10})] \\ - \frac{2(\mu_{10} - \mu_{00})}{\sqrt{(\mu_{11} - \mu_{10})(\mu_{00} - \mu_{10})}} \tan^{-1} \frac{(\mu_{11} - 2\mu_{10} + \mu_{00})p + (\mu_{10} - \mu_{00})}{\sqrt{(\mu_{11} - \mu_{10})(\mu_{00} - \mu_{10})}} \} \frac{p_n}{p_0}. \end{aligned} \quad (5.41)$$

We now take up the case in which there is no dominance. In this case we have $\mu_{11} - \mu_{10} = \mu_{10} - \mu_{00}$. Consequently, the solution given by (5.40) reduces to

$$\frac{(x_n - \mu_{10})^2}{2\sigma^2} - \frac{(x_0 - \mu_{10})^2}{2\sigma^2} = \ln(2p_n - 1) - \ln(2p_0 - 1) . \quad (5.42)$$

Using equation (5.42), we find

$$\begin{aligned} N_{10}(x_n) &= \frac{1}{\sigma\sqrt{2\pi}} \exp \left[- \frac{(x_n - \mu_{10})^2}{2\sigma^2} \right] \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp \left[- \frac{(x_0 - \mu_{10})^2}{2\sigma^2} + \ln \frac{2p_0 - 1}{2p_n - 1} \right] \\ &= \frac{1}{\sigma\sqrt{2\pi}} \frac{2p_0 - 1}{2p_n - 1} \exp \left[- \frac{(x_0 - \mu_{10})^2}{2\sigma^2} \right] . \end{aligned} \quad (5.43)$$

Replacing $N_{10}(x_n)$ in (5.19) by its value from (5.43), we find

$$dn = \frac{S \sigma\sqrt{2\pi} \exp[(x_0 - \mu_{10})^2/2\sigma^2]}{(\mu_{10} - \mu_{00})(2p_0 - 1)} \frac{2p_n - 1}{p_n(1 - p_n)} dp . \quad (5.44)$$

Equation (5.44) yields the following solution

$$n = \frac{S \sigma\sqrt{2\pi} \exp[(x_0 - \mu_{10})^2/2\sigma^2]}{(\mu_{10} - \mu_{00})(2p_0 - 1)} [\ln p_0(1 - p_0) - \ln p_n(1 - p_n)] . \quad (5.45)$$

If we let $\mu_{11} = 2$, $\mu_{10} = 1$ and $\mu_{00} = 0$, equation (5.45) becomes

$$n = \frac{S \sigma\sqrt{2\pi} \exp[(x_0 - 1)^2/2\sigma^2]}{2p_0 - 1} [\ln p_0(1 - p_0) - \ln p_n(1 - p_n)] . \quad (5.46)$$

If we let $\mu_{11} = u\sigma$, $\mu_{10} = 0$ and $\mu_{00} = -u\sigma$, equation (5.45) becomes

$$n = \frac{S \sqrt{2\pi} \exp(x_0^2/2\sigma^2)}{u(2p_0-1)} [\ln p_0(1-p_0) - \ln p_n(1-p_n)] . \quad (5.47)$$

Equation (5.46) may be written in the form

$$p_n = \frac{1}{2} \pm \frac{1}{2} \sqrt{1 + 4p_0(1-p_0) \exp \left[-\frac{(2p_0-1)n}{S\sigma\sqrt{2\pi} \exp[(x_0-1)^2/2\sigma^2]} \right]} . \quad (5.48)$$

For complete dominance model, i.e. $\mu_{11} = \mu_{10} > \mu_{00}$, equation (5.40) becomes

$$\frac{(x_n - \mu_{10})^2}{2\sigma^2} - \frac{(x_0 - \mu_{10})^2}{2\sigma^2} = 2[\ln(1-p_n) - \ln(1-p_0)] . \quad (5.49)$$

Using equation (5.49), we find

$$\begin{aligned} N_{10}(x_n) &= \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x_n - \mu_{10})^2}{2\sigma^2} \right] \\ &= \frac{1}{\sigma\sqrt{2\pi}} \frac{(1-p_0)^2}{(1-p_n)^2} \exp \left[-\frac{(x_0 - \mu_{10})^2}{2\sigma^2} \right] . \end{aligned} \quad (5.50)$$

Replacing $N_{10}(x_n)$ in (5.19) by its value from (5.50), we find

$$dn = \frac{S \sigma\sqrt{2\pi} \exp[(x_0 - \mu_{10})^2/2\sigma^2]}{(\mu_{10} - \mu_{00})(1-p_0)^2} \frac{1}{p} dp . \quad (5.51)$$

Equation (5.51) yields the following solution

$$n = \frac{S \sigma \sqrt{2\pi} \exp(x_0 - \mu_{10})^2 / 2\sigma^2}{(\mu_{10} - \mu_{00})(1-p_0)^2} [\ln p_n - \ln p_0] \quad (5.52)$$

The solution (5.52) may be written in the form

$$p_n = p_0 \exp \left[\frac{S \sigma \sqrt{2\pi} \exp[(x_0 - \mu_{10})^2 / 2\sigma^2]}{(\mu_{10} - \mu_{00})(1-p_0)^2} n \right] \quad (5.53)$$

If we let $\mu_{11} = \mu_{10} = k\sigma$ and $\mu_{00} = 0$, equation (5.52) becomes

$$n = \frac{S \sigma \sqrt{2\pi} \exp[(x_0 - k\sigma)^2 / 2\sigma^2]}{k(1-p_0)^2} [\ln p_n - \ln p_0] \quad (5.54)$$

If we let $\mu_{11} = \mu_{10} = 1$ and $\mu_{00} = 0$, equation (5.52) becomes

$$n = \frac{S \sigma \sqrt{2\pi} \exp[(x_0 - 1)^2 / 2\sigma^2]}{(1-p_0)^2} [\ln p_n - \ln p_0] \quad (5.55)$$

Griffing (1960) examined the theoretical consequences of culling selection by utilizing the same assumptions given in this section. Consider the three genotypes originated by one locus with two alleles, A_1 and A_0 , their mean phenotypic values and frequencies in generation n given in Table 21, where μ denotes the population mean.

Table 21. Mean phenotypic values and frequencies of genotypes.

	Genotype		
	A_1A_1	A_1A_0	A_0A_0
Mean phenotypic value	μ_{11}	μ_{10}	μ_{00}
Coded mean phenotypic value	$d_{11} = \mu_{11} - \mu$	$d_{10} = \mu_{10} - \mu$	$d_{00} = \mu_{00} - \mu$
Frequency	p_n^2	$2p_n(1-p_n)$	$(1-p_n)^2$

Following Griffing's approach and for the one locus with two alleles case, the frequency of A_1 allele in the selected population in generation n is

$$\begin{aligned}
 p_{n+1} &= p_n^2 + \frac{i}{\sigma^2} p_n^2 d_{11} + p_n(1-p_n) + \frac{i}{\sigma^2} p_n(1-p_n) d_{10} \\
 &= p_n + \frac{i}{\sigma^2} p_n [p_n d_{11} + (1-p_n) d_{10}] , \quad (5.56)
 \end{aligned}$$

where i is the selection differential (Griffing, 1960). Equation (5.56) for the gene frequency may be written in the form

$$p_{n+1} = p_n + \frac{\bar{i}}{\sigma} \alpha p_n , \quad (5.57)$$

where

$\bar{i} = \frac{i}{\sigma}$ is the standardized form of the selection differential

$\alpha = p_n d_{11} + (1-p_n) d_{10}$ is the additive effect of the A_1 allele.

This approach was followed by Latter (1965).

The change in gene frequency in one generation is given by

$$\Delta p_n = p_{n+1} - p_n .$$

Using (5.56), we find

$$\Delta p_n = \frac{\bar{i}}{\sigma} p_n [p_n d_{11} + (1-p_n)d_{10}] . \quad (5.58)$$

From Table 21 we have

$$\begin{aligned} d_{11} &= \mu_{11} - \mu = \mu_{11} - [\mu_{11}p^2 + \mu_{10}2p(1-p) + \mu_{00}(1-p)^2] \\ d_{10} &= \mu_{10} - \mu = \mu_{10} - [\mu_{11}p^2 + \mu_{10}2p(1-p) + \mu_{00}(1-p)^2] . \end{aligned} \quad (5.59)$$

Replacing d_{11} and d_{10} in (5.58) by their values from (5.59) we find

$$\Delta p_n = \frac{\bar{i}}{\sigma} p_n [(\mu_{10} - \mu_{00}) + (\mu_{11} - 3\mu_{10} + 2\mu_{00})p_n - (\mu_{11} - 2\mu_{10} + \mu_{00})p_n^2] . \quad (5.60)$$

With no dominance we have $\mu_{11} - \mu_{10} = \mu_{10} - \mu_{00} = \frac{k}{2}$, where $k = \mu_{11} - \mu_{00}$.

In this case, the expression for the change in gene frequency given by

(5.60) reduces to

$$\Delta p_n = \frac{\bar{i}}{\sigma} \frac{k}{2} p_n (1-p_n) . \quad (5.61)$$

Denote by σ_g^2 the genotypic variance and by σ_e^2 the environmental variance.

Then, if there is no genotype by environment interaction, the total phenotypic variance, σ^2 , may be written in the form

$$\sigma^2 = \sigma_g^2 + \sigma_e^2 \quad (5.62)$$

(Kempthorne, 1957). With additive gene effects we have $\sigma_g^2 = \sigma_a^2$, where σ_a^2 is the additive variance. In this case, the total phenotypic variance given by (5.62) may be written in the form

$$\sigma^2 = \sigma_a^2 + \sigma_e^2 . \quad (5.63)$$

With one locus and two alleles and no dominance we have

$$\sigma_a^2 = \frac{1}{2} k^2 p_n (1-p_n) ,$$

where $\frac{k}{2} = \mu_{11} - \mu_{10} = \mu_{10} - \mu_{00}$. Consequently, we have

$$\sigma^2 = \frac{1}{2} k^2 p_n (1-p_n) + \sigma_e^2 . \quad (5.64)$$

If there is no environmental variability, i.e. $\sigma_e^2 = 0$, the total phenotypic variance given by (5.64) becomes

$$\sigma^2 = \frac{1}{2} k^2 p_n (1-p_n) , \quad (5.65)$$

which changes markedly with changes in p_n . The change in gene frequency given by (5.61) may be continuized by writing

$$\begin{aligned} \Delta p_n &= \frac{\bar{i}}{\sigma} \frac{k}{2} p_n (1-p_n) \\ &= \frac{i}{\sigma^2} \frac{k}{2} p_n (1-p_n) \\ &\doteq \frac{dp}{dn} . \end{aligned} \quad (5.66)$$

If σ^2 is taken to be constant, equation (5.66) yields the following solution

$$\ln p_n - \ln(1-p_n) = \ln p_0 - \ln(1-p_0) + \frac{i}{\sigma^2} \frac{k}{2} n, \quad (5.67)$$

where p_0 is the frequency of A_1 in the initial population. Equation (5.67) shows that the function of the gene frequency given by the left-hand side of (5.67) is expected to change linearly with generations with slope given by $(ik/2\sigma^2)$ when σ^2 is constant. Equation (5.67) may be written in the form

$$p_n = \frac{1}{1 + \frac{1-p_0}{p_0} e^{-\frac{i k}{2\sigma^2} n}}. \quad (5.68)$$

It is seen from (5.68) that p approaches one as n approaches infinity.

We now consider the case in which σ^2 is not constant. With no dominance and no environmental variability the total phenotypic variance is given by (5.65) above. If we replace σ^2 in equation (5.66) by its value from (5.65), we find

$$\frac{dp}{dn} = \frac{i}{k}. \quad (5.69)$$

Equation (5.69) yields the following solution:

$$p_n = p_0 + \frac{i}{k} n. \quad (5.70)$$

The above solution (5.70) indicates that p_n is expected to change linearly with generations with slope given by (i/k) when $\sigma_e^2 = 0$ and σ^2 is not

constant. Table 22 gives the values of the slope (i/k) for different values of S when $k = 2$.

Table 22. Values of the slope (i/k) for different values of selection (S) when $k = 2$.

S	.95	.90	.80	.75	.50	.25
$\frac{i}{k} = \frac{i}{2}$.054	.097	.175	.211	.398	.635

Table 23 gives the number of generations required to change gene frequency from its initial value of .05 to .5 for different values of selection (S) and environmental variance (σ^2) for the case of no dominance. In the table, n is the number of generations obtained using the computer results, n_1 is the number obtained using equation (5.28), n_2 is the number using equation (5.46), and n_3 is the number obtained using equation (5.70).

Table 23. Number of generations required for gene frequency to increase from .05 to .5. No dominance.

			Environmental Variance σ^2				
			0	1	2	3	4
Selection	$S = .25$	n	2	3	4	5	5
		n_1	2	2	4	6	8
		n_2	2	3	4	6	8
		n_3	1	2	5	7	9
	$S = .50$	n	3	5	7	8	9
		n_1	3	4	8	11	15
		n_2	3	4	8	10	14
		n_3	2	4	8	11	15

Table 23 (continued)

			Environmental Variance σ^2				
			0	1	2	3	4
Selection	S = .75	n	8	11	13	15	16
		n_1	8	6	12	17	22
		n_2	8	8	13	18	23
		n_3	3	7	14	21	28

It is seen from the above table that, when $\sigma^2 = 0$, the approach given in the present study in the first section is more accurate for estimating the change in gene frequency than using Griffing's approach (1960). When $\sigma^2 > 0$, it is seen that Griffing's approach over estimates the change when selection is effective and under-estimates the change when selection is not effective. The approach given in this section gives intermediate results.

With complete dominance we have $\mu_{11} = \mu_{10} = v$, say, and $\mu_{00} = 0$. Consequently, equation (5.60) for the change in gene frequency redut to

$$\Delta p_n = \frac{i}{\sigma^2} v p_n (1-p_n)^2 . \quad (5.71)$$

The change in gene frequency may be continuized by writing

$$\Delta p_n = \frac{i}{\sigma^2} v p_n (1-p_n)^2 \doteq \frac{dp}{dn} . \quad (5.72)$$

Using equation (5.62), the total phenotypic variance for the present model is given by

$$\begin{aligned} \sigma^2 &= \sigma_g^2 + \sigma_e^2 \\ &= v^2 p(2-p)(1-p)^2 + \sigma_e^2 . \end{aligned} \quad (5.73)$$

For the case of no variability, i.e. $\sigma_e^2 = 0$, the total phenotypic variance for the present model reduces to

$$\sigma^2 = \sigma_g^2 = v^2 p(2-p)(1-p)^2 . \quad (5.74)$$

Replacing σ^2 in (5.72) by its value from (5.74) we find

$$\frac{dp}{dn} = \frac{i}{v(2-p_n)} . \quad (5.75)$$

Equation (5.75) yields the following solution

$$p_n(4-p_n) = p_0(4-p_0) + \frac{4i}{v} n . \quad (5.76)$$

The solution given by (5.76) shows that the function of gene frequency $p(4-p)$ is expected to change linearly with generations with slope given by $(2i/v)$ when $\sigma^2 = 0$.

If σ^2 is taken to be constant, equation (5.72) yields the following solution

$$\frac{1}{1-p_n} - \ln \frac{1-p_n}{p_n} = \frac{1}{1-p_0} - \ln \frac{1-p_0}{p_0} + \frac{i}{\sigma^2} v n . \quad (5.77)$$

The function of gene frequency given by the left-hand side of (5.77) is expected to change linearly with generations with slope given by (iv/σ^2) when σ^2 is constant. Table 24 gives the number of generations required to change gene frequency from its initial value of .05 to .5 for different values of selection (S) and environmental variability (σ^2) for the complete dominance model. In the table, n is the number of generations obtained using the computer results, n_1 is the number obtained using equation (5.35) n_2 is the number obtained using (5.55) and n_3 is the number obtained using (5.77).

Table 24. Number of generations required for gene frequency to change from .05 to .5. Complete dominance.

			Environmental Variance σ^2				
			0	1	2	3	4
Selection	S = .25	n	2	4	5	6	7
		n_1	2	3	5	7	10
		n_2	2	4	6	7	8
		n_3	1	4	7	10	13
	S = .50	n	4	6	8	10	11
		n_1	4	5	10	15	19
		n_2	4	6	9	14	20
		n_3	1	5	10	15	20
	S = .75	n	8	12	15	18	20
		n_1	8	8	15	22	29
		n_2	8	10	16	20	28
		n_3	2	10	19	28	37

For the complete dominance model, it is seen from the above table that the approach presented in the present study in the first section gives close estimates of the change in gene frequency when there is no environmental variability, while Griffing's approach over estimates the change. When $\sigma^2 > 0$, both n_1 and n_2 give closer estimate of n than n_3 . In general, n_1 , n_2 and n_3 over-estimate the change when selection is strong and under-estimate the change when selection is weak. Exact solution of the continuous analog taking account of non-zero σ_e^2 and the change in σ_g^2 could not be obtained.

Equilibrium of gene frequency exists when $\Delta p = 0$, or, using equation (5.17),

$$\frac{1}{S} p(1-p) N_{10}(x) [p(\mu_{11} - \mu_{10}) + (1-p)(\mu_{10} - \mu_{00})] = 0 .$$

There are four possible solutions

(i) $p = 0$

(ii) $p = 1$

(iii) $p = \frac{\mu_{00} - \mu_{10}}{\mu_{11} - 2\mu_{10} + \mu_{00}}$ (5.78)

(iv) $N_{10}(x) = 0$.

The first three solutions may also be obtained using Griffing's approach (equation 5.60). With regard to the stability of the above four equilibrium points, the first two conditions are trivially stable. With the complete dominance model, for which $\mu_{11} = \mu_{10} > \mu_{00}$, the second and the third conditions are the same. The third condition was given by Fisher (1930) in terms of constant selective advantages of the genotypes. For the third equilibrium point given in (5.78) to be stable it is seen that the heterozygotes must be superior over both homozygotes. If $\mu_{11} > \mu_{10} > \mu_{00}$, the population will become (1,0), and it is stable. Using scale 0, 1, 2 for the μ 's, the following diagram gives the equilibrium point for different models.

	0	1	2	Equilibrium point
Additive	μ_{00}	μ_{10}	μ_{11}	$p = 1$
Complete dominance	μ_{00}	$\mu_{10} = \mu_{11}$		$p = 1$
Overdominance	$\mu_{00} = \mu_{11}$	μ_{10}		$p = 1/2$

It is seen that in order to have a non-trivial point of stable equilibrium overdominance must be present. Kimura (1956) and Tallis (1966) show that in order for the non-trivial equilibrium to be stable the quantities $(\mu_{11}-\mu_{10})$ and $(\mu_{00}-\mu_{10})$ must have the same sign; that is, the heterozygote fitness must be either greater than or less than those of both homozygotes.

Following the approach given by Lewontin (1958) and Tallis (1966), an equilibrium point \hat{p} will be stable if $\frac{d\Delta p}{dp}$ evaluated at \hat{p} is negative. Using (5.17), we find

$$\begin{aligned} \frac{d\Delta p}{dp} = & \frac{1}{S} N_{10}(x_n) \{ [(\mu_{10}-\mu_{00}) + 2(\mu_{11}-3\mu_{10}+2\mu_{00})p_n - 3(\mu_{11}-2\mu_{10}+\mu_{00})p_n^2] \\ & - p_n(1-p_n) \frac{(x_n-\mu_{10})}{2\sigma^2} [(\mu_{10}-\mu_{00}) - (\mu_{11}-2\mu_{10}+\mu_{00})p_n] \frac{dx_n}{dn} \}. \end{aligned} \quad (5.79)$$

Using the equilibrium point given by the third condition in (5.78), equation (5.79) reduces to

$$\left. \frac{d\Delta p}{dp} \right|_{\hat{p}} = \frac{1}{S} N_{10}(x_n) \frac{(\mu_{11}-\mu_{10})(\mu_{00}-\mu_{10})}{\mu_{11} - 2\mu_{10} + \mu_{00}}. \quad (5.80)$$

For the complete dominance case the above expression (5.80) is zero.

Lewontin (1958) defines this case as a state of neutrality. It is seen from (5.80) that the sign will be negative when overdominance is present.

The fourth equilibrium condition given in (5.78) affects the approach to the equilibrium rather than its value. Robertson (1956) showed that when the amount of variability is the same for each genotype any equilibrium value for p between 0 and 1 is unstable equilibrium and one or other allele will eventually become fixed unless heterozygotes are superior over both homozygotes. It is seen from equation (5.17) for the change in gene

frequency that selection as well as the magnitude of the environmental variability affect the approach to the equilibria. When selection is effective, that is, when both S and σ are small, the approach to the equilibrium point will be faster.

Two-loci Case

We consider in this section the case of two loci A and B with two alleles each A, a and B, b . There are four possible gametic types AB, Ab, aB and ab with frequencies x_1, x_2, x_3 and x_4 , respectively. The recombination frequency between the two loci is denoted by r , with r being constant and the same in the two sexes. The zygotic and gametic frequencies for the ten possible genotypes are presented in Table 25.

Table 25. The zygotic and gametic frequencies and their survival probabilities.

Zygotes	Frequency	Gametic Frequencies				Mean pheno- typic value	Probability of survival
		AB	Ab	aB	ab		
$AABB$	x_1^2	1	0	0	0	μ_{22}	M_{22}
$AaBB$	$2x_1x_3$	$\frac{1}{2}$	0	$\frac{1}{2}$	0	μ_{12}	M_{12}
$AABb$	$2x_1x_2$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	μ_{21}	M_{21}
AB/ab	$2x_1x_4$	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$	μ_{11}	M_{11}
Ab/aB	$2x_2x_3$	$\frac{1}{2}$	$\frac{1}{2}(1-r)$	$\frac{1}{2}(1-r)$	$\frac{1}{2}$	μ_{11}	M_{11}
$AAbb$	x_2^2	0	1	0	0	μ_{20}	M_{20}
$Aabb$	$2x_2x_4$	0	$\frac{1}{2}$	0	$\frac{1}{2}$	μ_{10}	M_{10}
$aaBB$	x_3^2	0	0	1	0	μ_{02}	M_{02}
$aaBb$	$2x_3x_4$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	μ_{01}	M_{01}
$aabb$	x_4^2	0	0	0	1	μ_{00}	M_{00}

Gametic disequilibrium will be denoted by D , where $D = x_1x_4 - x_2x_3$. Following the one locus case given in the previous section, we assume that the genotypes are variable so that the value on the scale of merit of an individual with genotype $A_iA_jB_kB_l$ is normally distributed around some fixed value μ_{ijkl} with variance σ_{ijkl}^2 , independently from individual to individual. We assume that the population is very large so that sampling element due to finite size can be ignored. In any generation, the individuals which are in the top fraction S ($0 < S < 1$) of the whole population of phenotypic values are saved to reproduce the next generation through random mating. The fraction S is assumed to be constant and the same in the two sexes. We assume that there are no differences in the fecundity of matings. We now proceed to examine the progress of the population with a given set of parameters S , μ_{ijkl} , σ_{ijkl}^2 , and r as well as the relationships between the effects of these parameters.

For the case of 2 alleles at each locus the probability that an individual of a given genotype survives selection in generation n can be written as

$$\text{Prob}(t \geq T_n) = \frac{1}{\sigma_{ij}\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{ij})^2}{2\sigma_{ij}^2} \right] dt, \quad (6.1)$$

where t is the variate, T_n is the truncation point in generation n , and the subscripts i and j denote the number of A and B genes present in the genotype, respectively. The proportion of individuals selected in generation n is

$$\begin{aligned}
S = & x_{1,n}^2 \frac{1}{\sigma_{22}\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{22})^2}{2\sigma_{22}^2} \right] dt \\
& + 2x_{1,n}x_{2,n} \frac{1}{\sigma_{21}\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{21})^2}{2\sigma_{21}^2} \right] dt \\
& + 2x_{1,n}x_{3,n} \frac{1}{\sigma_{12}\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{12})^2}{2\sigma_{12}^2} \right] dt \\
& + 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) \frac{1}{\sigma_{11}\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{11})^2}{2\sigma_{11}^2} \right] dt \\
& + x_{2,n}^2 \frac{1}{\sigma_{20}\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{20})^2}{2\sigma_{20}^2} \right] dt \\
& + x_{3,n}^2 \frac{1}{\sigma_{02}\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{02})^2}{2\sigma_{02}^2} \right] dt \\
& + 2x_{2,n}x_{4,n} \frac{1}{\sigma_{10}\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{10})^2}{2\sigma_{10}^2} \right] dt \\
& + 2x_{3,n}x_{4,n} \frac{1}{\sigma_{01}\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{01})^2}{2\sigma_{01}^2} \right] dt \\
& + x_{4,n}^2 \frac{1}{\sigma_{00}\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{00})^2}{2\sigma_{00}^2} \right] dt . \tag{6.2}
\end{aligned}$$

We now take up the case in which $\sigma_{ij}^2 = \sigma^2$ is the same for all genotypes.

Let

$$N_{ij}(t_n) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{(t_n - \mu_{ij})^2}{2\sigma^2} \right] \quad (6.3)$$

and

$$\begin{aligned} M_{ij}(T_n) &= \frac{1}{\sigma \sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t - \mu_{ij})^2}{2\sigma^2} \right] dt \\ &= \int_{T_n}^{\infty} N_{ij}(t) dt . \end{aligned} \quad (6.4)$$

Equation (6.2) may now be written in the form

$$\begin{aligned} S &= x_{1,n}^2 M_{22}(T_n) + 2x_{1,n}x_{2,n} M_{21}(T_n) + 2x_{1,n}x_{3,n} M_{12}(T_n) \\ &\quad + 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) M_{11}(T_n) + x_{2,n}^2 M_{20}(T_n) \\ &\quad + x_{3,n}^2 M_{02}(T_n) + 2x_{2,n}x_{4,n} M_{10}(T_n) \\ &\quad + 2x_{3,n}x_{4,n} M_{01}(T_n) + x_{4,n}^2 M_{00}(T_n) \end{aligned} \quad (6.5)$$

Using a Taylor's series expansion around μ_{11} we find

$$M_{ij}(T_n) = M_{11}(T_n) + (\mu_{ij} - \mu_{11}) M'_{11} + \frac{1}{2} (\mu_{ij} - \mu_{11})^2 M''_{11} + \dots \quad (6.6)$$

where the primes denote the degree of the derivative. If the differences between the μ_{ij} 's are small so that orders higher than the first can be ignored, equation (6.6) can be written approximately in the form

$$M_{ij}(T_n) = M_{11}(T_n) + (\mu_{ij} - \mu_{11}) M'_{11} \quad i = 0,1; j = 0,1 . \quad (6.7)$$

Differentiating $M_{11}(T_n)$ within the integral with respect to μ_{11} we find

$$\begin{aligned}
 M'_{11}(T_n) &= \frac{d}{d\mu_{11}} M_{11}(T_n) \\
 &= \frac{d}{d\mu_{11}} \frac{1}{\sigma\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{11})^2}{2\sigma^2} \right] dt \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \int_{T_n}^{\infty} \frac{(t-\mu_{11})}{\sigma^2} \exp \left[-\frac{(t-\mu_{11})^2}{2\sigma^2} \right] dt \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(T_n-\mu_{11})^2}{2\sigma^2} \right] \\
 &= N_{11}(T_n) , \tag{6.8}
 \end{aligned}$$

where $N_{11}(T_n)$ is defined in (6.3). Differentiating $M_{11}(T_n)$ with respect to n we find

$$\frac{d}{dn} M_{11}(T_n) = \frac{dT_n}{dn} \frac{d}{dT_n} M_{11}(T_n) = - N_{11}(T_n) \frac{dT_n}{dn} . \tag{6.9}$$

Differentiating $N_{11}(T_n)$ defined in (6.3) with respect to n we find

$$\begin{aligned}
 \frac{d}{dn} N_{11}(T_n) &= -\frac{(T_n-\mu_{11})}{\sigma^2} \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(T_n-\mu_{11})^2}{2\sigma^2} \right] \frac{dT_n}{dn} \\
 &= -\frac{(T_n-\mu_{11})}{\sigma^2} N_{11}(T_n) \frac{dT_n}{dn} . \tag{6.10}
 \end{aligned}$$

Using equations (6.7) and (6.8), equation (6.5) may be written in the approximate form

⋮

$$\begin{aligned}
S = & [M_{11}(T_n) + (\mu_{22} - \mu_{11}) N_{11}(T_n)] x_{1,n}^2 \\
& + [M_{11}(T_n) + (\mu_{21} - \mu_{11}) N_{11}(T_n)] 2x_{1,n}x_{2,n} \\
& + [M_{11}(T_n) + (\mu_{12} - \mu_{11}) N_{11}(T_n)] 2x_{1,n}x_{3,n} \\
& + [M_{11}(T_n)] 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) \\
& + [M_{11}(T_n) + (\mu_{20} - \mu_{11}) N_{11}(T_n)] x_{2,n}^2 \\
& + [M_{11}(T_n) + (\mu_{02} - \mu_{11}) N_{11}(T_n)] x_{3,n}^2 \\
& + [M_{11}(T_n) + (\mu_{10} - \mu_{11}) N_{11}(T_n)] 2x_{2,n}x_{4,n} \\
& + [M_{11}(T_n) + (\mu_{01} - \mu_{11}) N_{11}(T_n)] 2x_{3,n}x_{4,n} \\
& + [M_{11}(T_n) + (\mu_{00} - \mu_{11}) N_{11}(T_n)] x_{4,n}^2 \\
= & M_{11}(T_n) + N_{11}(T_n) f_n(\underline{x}, \underline{\mu}) , \tag{6.11}
\end{aligned}$$

where

$$\begin{aligned}
f_n(\underline{x}, \underline{\mu}) = & (\mu_{22} - \mu_{11}) x_{1,n}^2 + (\mu_{21} - \mu_{11}) 2x_{1,n}x_{2,n} \\
& + (\mu_{12} - \mu_{11}) 2x_{1,n}x_{3,n} + (\mu_{20} - \mu_{11}) x_{2,n}^2 \\
& + (\mu_{02} - \mu_{11}) x_{3,n}^2 + (\mu_{10} - \mu_{11}) 2x_{2,n}x_{4,n} \\
& + (\mu_{01} - \mu_{11}) 2x_{3,n}x_{4,n} + (\mu_{00} - \mu_{11}) x_{4,n}^2 \tag{6.12}
\end{aligned}$$

Because S is constant in every generation, we have

$$\frac{d}{dn} S = 0 .$$

Consequently, differentiating S given by (6.11) with respect to n we find

$$\begin{aligned} 0 &= \frac{d}{dn} S = \frac{d}{dn} [M_{11}(T_n) + N_{11}(T_n) f_n(\underline{x}, \underline{\mu})] \\ &= M'_{11}(T_n) + N'_{11}(T_n) f_n(\underline{x}, \underline{\mu}) + N_{11}(T_n) f'_n(\underline{x}, \underline{\mu}), \end{aligned} \quad (6.13)$$

where $f_n(\underline{x}, \underline{\mu})$ is defined in (6.12). Using equations (6.9) and (6.10), equation (6.13) can be written as

$$\begin{aligned} 0 &= -N_{11}(T_n) \frac{dT_n}{dn} + N_{11}(T_n) [f'_n(\underline{x}, \underline{\mu}) \frac{d\bar{\underline{x}}}{dn}] \\ &\quad - \frac{(T_n - \mu_{11})}{\sigma^2} N_{11}(T_n) \frac{dT_n}{dn} f_n(\underline{x}, \underline{\mu}), \end{aligned} \quad (6.14)$$

where $f_n(\underline{x}, \underline{\mu})$ is defined in (6.12) and

$$\begin{aligned} [f'_n(\underline{x}, \underline{\mu}) \frac{d\bar{\underline{x}}}{dn}] &= (\mu_{22} - \mu_{11}) 2x_{1,n} \frac{dx_1}{dn} + 2(\mu_{21} - \mu_{11})(x_{1,n} \frac{dx_2}{dn} + x_{2,n} \frac{dx_1}{dn}) \\ &\quad + 2(\mu_{12} - \mu_{11})(x_{1,n} \frac{dx_3}{dn} + x_{3,n} \frac{dx_1}{dn}) + 2(\mu_{20} - \mu_{11})x_{2,n} \frac{dx_2}{dn} \\ &\quad + 2(\mu_{02} - \mu_{11})x_{3,n} \frac{dx_3}{dn} + 2(\mu_{10} - \mu_{11})(x_{2,n} \frac{dx_4}{dn} + x_{4,n} \frac{dx_2}{dn}) \\ &\quad + 2(\mu_{01} - \mu_{11})(x_{3,n} \frac{dx_4}{dn} + x_{4,n} \frac{dx_3}{dn}) + 2(\mu_{00} - \mu_{11})x_{4,n} \frac{dx_4}{dn}. \end{aligned} \quad (6.15)$$

Equation (6.14) defines T_n in terms of the gametic frequencies $x_{i,n}$ ($i = 1, 2, 3, 4$) and vice versa. Equation (6.14) may be written symbolically in the form

$$[1 + \frac{(T_n - \mu_{11})}{\sigma^2} f_n(\underline{x}, \underline{\mu})] dT_n = [f'_n(\underline{x}, \underline{\mu}) d\bar{\underline{x}}], \quad (6.16)$$

where $f_n(\underline{x}, \underline{\mu})$ and $[f'_n(\underline{x}, \underline{\mu}) d\bar{x}]$ are defined in (6.12) and (6.15), respectively. A solution to the non-linear differential equation (6.16) is very difficult to obtain mathematically and one may solve the equation numerically on the computer. However, two extreme cases may be considered.

Case 1: σ is very large.

In this case the term $(T_n - \mu_{11})/\sigma^2$ is assumed to be very small. Consequently, equation (6.16) may be written symbolically as

$$dT_n = [f'_n(\underline{x}, \underline{\mu}) d\bar{x}], \quad (6.17)$$

where $[f'_n(\underline{x}, \underline{\mu}) d\bar{x}]$ is defined in (6.15). Equation (6.17) yields the following solution

$$T_n - T_0 = f_n(\underline{x}, \underline{\mu}) - f_0(\underline{x}, \underline{\mu}), \quad (6.18)$$

where T_0 and $f_0(\underline{x}, \underline{\mu})$ are the truncation point and frequencies in the initial population, respectively. Because T_n , the truncation point in generation n , is related to fitness defined by M_{ij} in (6.4), it is seen from (6.18) that fitness depends on the initial structure of the population, the amount of variability, and the gametic frequencies in any given generation. Work with experimental populations has shown some existing evidence of such relation (Dobzhansky, 1947; Thomson, 1961; Kojima and Yarbrough, 1967).

The gametic frequencies after one generation of selection are given by

$$\begin{aligned}
x_{1,n+1} &= \frac{1}{S} \{x_{1,n}(x_{1,n}M_{22} + x_{2,n}M_{21} + x_{3,n}M_{12} + x_{4,n}M_{11}) - rD_n M_{11}\} \\
x_{2,n+1} &= \frac{1}{S} \{x_{2,n}(x_{1,n}M_{21} + x_{2,n}M_{20} + x_{3,n}M_{11} + x_{4,n}M_{10}) + rD_n M_{11}\} \\
x_{3,n+1} &= \frac{1}{S} \{x_{3,n}(x_{1,n}M_{12} + x_{2,n}M_{11} + x_{3,n}M_{02} + x_{4,n}M_{01}) + rD_n M_{11}\} \\
x_{4,n+1} &= \frac{1}{S} \{x_{4,n}(x_{1,n}M_{11} + x_{2,n}M_{10} + x_{3,n}M_{01} + x_{4,n}M_{00}) - rD_n M_{11}\} ,
\end{aligned} \tag{6.19}$$

where

$$D_n = x_{1,n}x_{4,n} - x_{2,n}x_{3,n} \tag{6.20}$$

is the gametic disequilibrium in generation n . Equations (6.19) are similar to those derived by Lewontin and Kojima (1960). As in their problem the frequencies x_1 , x_2 , x_3 and x_4 enter into the equations in a symmetrical manner.

Using equations (6.7) and (6.8), equations (6.19) may be written in the approximate form

$$\begin{aligned}
x_{1,n+1} &= \frac{1}{S} \{x_{1,n} [M_{11} + N_{11}(T_n)(x_{1,n}(\mu_{22}-\mu_{11}) + x_{2,n}(\mu_{21}-\mu_{11}) \\
&\quad + (\mu_{12}-\mu_{11})x_{3,n})] - rD_n M_{11}\}
\end{aligned}$$

$$\begin{aligned}
x_{2,n+1} &= \frac{1}{S} \{x_{2,n} [M_{11} + N_{11}(T_n)(x_{1,n}(\mu_{21}-\mu_{11}) + x_{2,n}(\mu_{20}-\mu_{11}) \\
&\quad + x_{4,n}(\mu_{10}-\mu_{11}))] + rD_n M_{11}\} \\
x_{3,n+1} &= \frac{1}{S} \{x_{3,n} [M_{11} + N_{11}(T_n)(x_{1,n}(\mu_{12}-\mu_{11}) + x_{3,n}(\mu_{02}-\mu_{11}) \\
&\quad + x_{4,n}(\mu_{01}-\mu_{11}))] + rD_n M_{11}\} \\
x_{4,n+1} &= \frac{1}{S} \{x_{4,n} [M_{11} + N_{11}(T_n)(x_{2,n}(\mu_{10}-\mu_{11}) + x_{3,n}(\mu_{01}-\mu_{11}) \\
&\quad + x_{4,n}(\mu_{00}-\mu_{11}))] - rD_n M_{11}\} . \tag{6.21}
\end{aligned}$$

Equations (6.21) show the dependence of the gametic frequencies in successive generations on selection, linkage, the amount of variability and on the spacing between the means of the genotypic classes. Because equations (6.21) are difficult to handle in the general form, some special cases of gene actions and interactions will be considered separately.

No dominance at both loci

The mean phenotypic values for the case of no dominance are presented in Table 27, Appendix B. With μ_{00} coded as 0, μ_{10} and μ_{01} as unity, the values of μ_{11} , μ_{20} , μ_{02} are 2, of μ_{21} and μ_{12} are 3, and of μ_{22} is 4. Accordingly, equation (6.11) takes the form:

$$S = M_{11}(T_n) + \mu_{11}(x_{1,n} - x_{4,n}) N_{11}(T_n) , \tag{6.22}$$

and if we use the mean phenotypic values given in Table 27, Appendix B, we find

$$S = M_{11}(T_n) + 2(x_{1,n} - x_{4,n}) N_{11}(T_n) . \tag{6.23}$$

Equations (6.21) for the new gametic frequencies may now be written in the form

$$\begin{aligned}
 x_{1,n+1} &= \frac{1}{S} \{x_{1,n} [M_{11} + \frac{1}{2} \mu_{11}(1 + x_{1,n} - x_{4,n})N_{11}(T_n)] - rD_n M_{11}\} \\
 x_{2,n+1} &= \frac{1}{S} \{x_{2,n} [M_{11} + \frac{1}{2} \mu_{11}(x_{1,n} - x_{4,n})N_{11}(T_n)] + rD_n M_{11}\} \\
 x_{3,n+1} &= \frac{1}{S} \{x_{3,n} [M_{11} + \frac{1}{2} \mu_{11}(x_{1,n} - x_{4,n})N_{11}(T_n)] + rD_n M_{11}\} \\
 x_{4,n+1} &= \frac{1}{S} \{x_{4,n} [M_{11} - \frac{1}{2} \mu_{11}(1 + x_{4,n} - x_{1,n})N_{11}(T_n)] - rD_n M_{11}\} .
 \end{aligned} \tag{6.24}$$

The change in the gametic frequencies in one generation is given by

$$\Delta x_{i,n} = x_{i,n+1} - x_{i,n} \quad i = 1, 2, 3, 4 .$$

Using equation (6.23) and (6.24) we find

$$\begin{aligned}
 \Delta x_{1,n} &= \frac{1}{S} \{ \frac{1}{2} x_{1,n} (1 - x_{1,n} + x_{4,n}) \mu_{11} N_{11}(T_n) - rD_n M_{11} \} \\
 \Delta x_{2,n} &= \frac{1}{S} \{ \frac{1}{2} x_{2,n} (x_{4,n} - x_{1,n}) \mu_{11} N_{11}(T_n) + rD_n M_{11} \} \\
 \Delta x_{3,n} &= \frac{1}{S} \{ \frac{1}{2} x_{3,n} (x_{4,n} - x_{1,n}) \mu_{11} N_{11}(T_n) + rD_n M_{11} \} \\
 \Delta x_{4,n} &= \frac{1}{S} \{ \frac{1}{2} x_{4,n} (x_{4,n} - x_{1,n} - 1) \mu_{11} N_{11}(T_n) - rD_n M_{11} \}
 \end{aligned} \tag{6.25}$$

The change in the gametic frequency may be continuized by writing

$$\frac{dx_{i,n}}{dn} \doteq \Delta x_{i,n} \quad i = 1, 2, 3, 4 . \tag{6.26}$$

The gametic disequilibrium in the (n+1)th generation is given by

$$\begin{aligned}
D_{n+1} &= x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1} \\
&= (x_{1,n} + \frac{dx_{1,n}}{dn})(x_{4,n} + \frac{dx_{4,n}}{dn}) - (x_{2,n} + \frac{dx_{2,n}}{dn})(x_{3,n} + \frac{dx_{3,n}}{dn}) \\
&\doteq D_n + x_{1,n} \frac{dx_{4,n}}{dn} + x_{4,n} \frac{dx_{1,n}}{dn} - x_{2,n} \frac{dx_{3,n}}{dn} - x_{3,n} \frac{dx_{2,n}}{dn} .
\end{aligned}$$

Using equations (6.25) and (6.26), we find, approximately,

$$\begin{aligned}
D_{n+1} &= D_n - \frac{1}{S} [(x_{1,n} - x_{4,n})\mu_{11}N_{11}(T_n) - rM_{11}] D_n \\
&= \frac{1}{S} \{S - [(x_{1,n} - x_{4,n})\mu_{11}N_{11}(T_n) - rM_{11}]\} D_n .
\end{aligned}$$

Substituting for S by its value from (6.22), we find

$$D_{n+1} = \frac{1}{S} (1+r) M_{11} D_n . \quad (6.27)$$

Equation (6.27) for the gametic disequilibrium indicates that if $D_0 = 0$, the population will remain in constant gametic equilibrium and no gametic disequilibrium will be built up by selection. The equation also indicates that selection will increase the amount of gametic disequilibrium. If we write

$$\begin{aligned}
M_{11}(T_n) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{11})^2}{2\sigma^2} \right] dt \\
&= \frac{1}{\sqrt{2\pi}} \int_{\frac{T_n - \mu_{11}}{\sigma}}^{\infty} \exp \left[-\frac{z^2}{2} \right] dz ,
\end{aligned}$$

where $z = (t - \mu_{11})/\sigma$, it is seen from (6.27) that gametic disequilibrium will decrease with the amount of variability. This is expected on the basis that selection will be more effective when the amount of variability is small. If $D_0 \neq 0$, gametic disequilibrium will diminish if the absolute value of the ratio D_{n+1}/D_n is less than one. Using equation (6.27), we find

$$\frac{D_{n+1}}{D_n} = \frac{1}{S} (1+r) M_{11}(T_n) ,$$

and the absolute value of this ratio will be less than one if $r < (S/M_{11}(T_n)) - 1$.

Under the assumption of the first case that σ is very large, so that the quantity $(T_n - \mu_{11})/\sigma^2$ is very small, we have, approximately,

$$\begin{aligned} N_{11}(T_n) &= \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(T_n - \mu_{11})^2}{2\sigma^2} \right] \\ &\doteq \frac{1}{\sigma\sqrt{2\pi}} . \end{aligned} \quad (6.28)$$

Accordingly, the change in the frequency x_1 in one generation given by (6.25) is approximately

$$\frac{dx_{1,n}}{dn} = \frac{1}{S} \left\{ \frac{1}{2\sigma\sqrt{2\pi}} x_{1,n} (1 - x_{1,n} + x_{4,n}) \mu_{11} - r D_n M_{11} \right\} . \quad (6.29)$$

If either D or r is zero and x_4 is small, equation (6.29) may be written in the approximate form

$$\frac{dx_{1,n}}{dn} = \frac{1}{2S\sigma\sqrt{2\pi}} x_{1,n} (1 - x_{1,n}) . \quad (6.30)$$

Equation (6.30) yields the following solution

$$x_{1,n} = \frac{1}{1 + \frac{1 - x_{1,0}}{x_{1,0}} e^{-\frac{\mu_{11}}{2S\sigma\sqrt{2\pi}} n}} \quad (6.31)$$

The solution (6.31) may also be written in the form

$$\ln(x_{1,n}) - \ln(1-x_{1,n}) = \ln(x_{1,0}) - \ln(1-x_{1,0}) + \frac{1}{2S\sigma\sqrt{2\pi}} n \quad (6.32)$$

Equation (6.32) indicates that the function of the gametic frequency x_1 given by the left-hand side is expected to change linearly with generations with slope given by

$$\hat{\beta} = 1/2S\sigma\sqrt{2\pi} \quad (6.33)$$

The slope $\hat{\beta}$ increases when selection is effective; that is, when both S and σ are small. Figure 8 shows the exact change in the gametic frequency x_1 with generations under different levels of selection, linkage and environmental variability when $x_{1,0} = .01$, $x_{4,0} = .81$, and $D_0 = 0$. It is seen from the figure that linkage has very little effect on the change in the gametic frequencies when selection is weak. Figure 9 shows the exact change in the function of the gametic frequency x_1 given by (6.32) with generations under different levels of selection, linkage and environmental variability when $x_{1,0} = .01$, $\mu_{11} = 2$ and $D_0 = 0$. Table 30, Appendix B, gives the exact slope β of the linear relation given by (6.32) and their estimates $\hat{\beta}$ given by (6.33) for different values of S , r and σ^2 . The values of β are obtained from Figure 9. It is seen from the table that $\hat{\beta}$ under-estimates β for all parameter values considered. Table 31, Appendix B,

gives the values of the exact slope β of the linear relation $\ln x_1$ with generations and their estimates $\hat{\beta}(= \frac{1-S}{S})$ for different values of S , r and σ^2 . This function of the gametic frequency $\ln x_1$ is the same function obtained under the case of two loci without variability (equation 3.25). It is seen from Table 31 that $\hat{\beta}$ give closer estimates for β during the early change in progress when selection is weak and linkage is tight. It is seen from Tables 30 and 31, Appendix B, that the magnitude of the amount of variability present and selection are important in determining the rate of change in gametic frequency.

Equilibrium in the gametic frequencies exists when $\Delta x_{i,n} = 0$ for all i ($= 1, 2, 3, 4$). Using equations (6.25), the equilibrium in gametic frequencies requires that

$$\begin{aligned}
 \frac{1}{2} x_{1,n} (1 - x_{1,n} + x_{4,n}) \mu_{11} N_{11}(T_n) - r D_n M_{11} &= 0 \\
 \frac{1}{2} x_{2,n} (x_{4,n} - x_{1,n}) \mu_{11} N_{11}(T_n) + r D_n M_{11} &= 0 \\
 \frac{1}{2} x_{3,n} (x_{4,n} - x_{1,n}) \mu_{11} N_{11}(T_n) + r D_n M_{11} &= 0 \\
 \frac{1}{2} x_{4,n} (x_{4,n} - x_{1,n} - 1) \mu_{11} N_{11}(T_n) - r D_n M_{11} &= 0 .
 \end{aligned} \tag{6.34}$$

If $D = 0$ at gametic equilibrium (Lewontin and Kojima, 1960), it is seen that equations (6.34) will be satisfied by the equilibrium frequencies $x_1 = 1$, $x_2 = x_3 = x_4 = 0$. At these equilibrium values the gametic disequilibrium will be zero. The quantity $N_{11}(T_n)$ appearing in (6.34) will affect the rate of approach to the equilibrium values.

The frequency of A gene in the (n+1)th generation is given by

$$p_{n+1} = x_{1,n+1} + x_{2,n+1} .$$

Using equations (6.24), we find

$$p_{n+1} = \frac{1}{S} \{ x_{1,n} [M_{11} + \frac{1}{2} \mu_{11} (1 + x_{1,n} - x_{4,n}) N_{11}(T_n)] \\ + x_{2,n} [M_{11} + \frac{1}{2} \mu_{11} (x_{1,n} - x_{4,n}) N_{11}(T_n)] \} .$$

Using equation (6.22), we find

$$p_{n+1} = p_n + \frac{1}{2S} \mu_{11} N_{11}(T_n) [x_{1,n} - p_n (x_{1,n} - x_{4,n})] . \quad (6.35)$$

The change in gene frequency after one generation is given by

$$\Delta p_n = p_{n+1} - p_n \\ = \frac{1}{2S} \mu_{11} N_{11}(T_n) [x_{1,n} - p_n (x_{1,n} - x_{4,n})] . \quad (6.36)$$

It is seen from (6.36) that equilibrium for gene frequency exists when $\Delta p_n = 0$, or when $p_n = x_1/(x_1 - x_4)$. The stable equilibrium value of p will be satisfied when $x_1 = 1$ and $x_4 = 0$. The effect of selection and environmental variability on the rate of approach to the equilibrium values is measured by S and $N_{11}(T_n)$ appearing in (6.36).

Complete dominance at both loci

The mean phenotypic values of the genotypes are given in Table 27, Appendix B. With μ_{00} coded as 0, the values of μ_{10} , μ_{01} , μ_{20} and μ_{02} coded as unity, the values of μ_{21} , μ_{12} , μ_{11} and μ_{22} are 2. Accordingly, equation (6.11) takes the form

$$S = M_{11} - \mu_{11} N_{11}(T_n) [x_{4,n}(1-x_{1,n}) + \frac{1}{2} (x_{2,n}^2 + x_{3,n}^2)] , \quad (6.37)$$

and if we use the mean phenotypic values given in Table 27, Appendix B, we find

$$S = M_{11} - 2N_{11}(T_n) [x_{4,n}(1-x_{1,n}) + \frac{1}{2} (x_{2,n}^2 + x_{3,n}^2)] . \quad (6.38)$$

Equations (6.21) for the new gametic frequencies may now be written in the form

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} (x_{1,n} - rD_n) M_{11} \\ x_{2,n+1} &= \frac{1}{S} \{x_{2,n} [M_{11} - \frac{1}{2} (x_{2,n} + x_{4,n})\mu_{11}N_{11}(T_n)] + rD_n M_{11}\} \\ x_{3,n+1} &= \frac{1}{S} \{x_{3,n} [M_{11} - \frac{1}{2} (x_{3,n} + x_{4,n})\mu_{11}N_{11}(T_n)] + rD_n M_{11}\} \\ x_{4,n+1} &= \frac{1}{S} \{x_{4,n} [M_{11} - \frac{1}{2} (1 + x_{4,n} - x_{1,n})\mu_{11}N_{11}(T_n)] - rD_n M_{11}\} . \end{aligned} \quad (6.39)$$

The change in the gametic frequencies in one generation is given by

$$\Delta x_{i,n} = x_{i,n+1} - x_{i,n} \quad i = 1, 2, 3, 4 .$$

Using equations (6.39), we find

$$\begin{aligned} \Delta x_{1,n} &= \frac{1}{S} \{ \frac{1}{2} x_{1,n} [2x_{4,n}(1-x_{1,n}) + (x_{2,n}^2 + x_{3,n}^2)]\mu_{11}N_{11}(T_n) - rD_n M_{11} \} \\ \Delta x_{2,n} &= \frac{1}{S} \{ \frac{1}{2} x_{2,n} [x_{4,n}(1-2x_{1,n}) - x_{2,n}(1-x_{2,n}) + x_{2,n}^2]\mu_{11}N_{11}(T_n) \\ &\quad + rD_n M_{11} \} \\ \Delta x_{3,n} &= \frac{1}{S} \{ \frac{1}{2} x_{3,n} [x_{4,n}(1-2x_{1,n}) - x_{3,n}(1-x_{3,n}) + x_{3,n}^2]\mu_{11}N_{11}(T_n) \\ &\quad + rD_n M_{11} \} \end{aligned} \quad (6.40)$$

$$\Delta x_{4,n} = \frac{1}{S} \left\{ \frac{1}{2} x_{4,n} [(x_{2,n}^2 + x_{3,n}^2) - x_{1,n} x_{4,n} - (1-x_{1,n})(1-x_{4,n})] \mu_{11} N_{11}(T_n) - r D_n M_{11} \right\} .$$

The gametic disequilibrium in the (n+1)th generation is given by

$$\begin{aligned} D_{n+1} &= x_{1,n+1} x_{4,n+1} - x_{2,n+1} x_{3,n+1} \\ &= (x_{1,n} + \frac{dx_{1,n}}{dn})(x_{4,n} + \frac{dx_{4,n}}{dn}) - (x_{2,n} + \frac{dx_{2,n}}{dn})(x_{3,n} + \frac{dx_{3,n}}{dn}) \\ &= D_n + x_{1,n} \frac{dx_{4,n}}{dn} + x_{4,n} \frac{dx_{1,n}}{dn} + \frac{dx_{1,n}}{dn} \frac{dx_{4,n}}{dn} - x_{2,n} \frac{dx_{3,n}}{dn} \\ &\quad - x_{3,n} \frac{dx_{2,n}}{dn} - (\frac{dx_{2,n}}{dn}) (\frac{dx_{3,n}}{dn}) . \end{aligned}$$

The change in the gametic frequency may be continuized by writing

$$\Delta x_{i,n} \doteq \frac{dx_{i,n}}{dn} \quad i = 1, 2, 3, 4 . \quad (6.41)$$

Using equations (6.41) and (6.40), the gametic disequilibrium in the (n+1)th generation is given approximately by

$$\begin{aligned} D_{n+1} &= D_n + \frac{1}{S} D_n \{ \mu_{11} N_{11}(T_n) [x_{4,n}(1-x_{1,n}) - \frac{1}{2}(1-x_{4,n}) \\ &\quad + (x_{2,n}^2 + x_{3,n}^2)] - r M_{11} \} . \end{aligned} \quad (6.42)$$

Using equation (6.37), equation (6.42) may be written in the form

$$D_{n+1} = \frac{1}{S} D_n \{ (1-r) M_{11} - \frac{1}{2} [(1-x_{4,n}) + (x_{2,n}^2 + x_{3,n}^2)] \} . \quad (6.43)$$

It is seen from equation (6.43) that when $D_0 = 0$ the population will remain in constant gametic equilibrium as for the case of two loci without variability (equation 3.61).

Consider the change in the frequency of x_1 given by the first equation in (6.40). Using equation (6.28), we have approximately,

$$\frac{dx_{1,n}}{dn} = \frac{1}{S} \left\{ \frac{1}{2} x_{1,n} [2x_{4,n}(1-x_{1,n}) + x_{2,n}^2 + x_{3,n}^2] \frac{\mu_{11}}{\sigma\sqrt{2\pi}} - rD_n M_{11} \right\}. \quad (6.44)$$

A solution to the above differential equation (6.44) is difficult to obtain without making further approximations. The two functions of x_1 obtained for the case of two loci with no variability, $\ln x_1$ with slope $\hat{\beta} = (1-S)/S$ (equation 3.55) and $\frac{1}{1-x_1} - \ln(1-x_1)$ with slope $\hat{\beta} = 1$ (equation 3.62), were used to linearize the change in the gametic frequency x_1 under the present model. Figure 8 shows the exact change in the gametic frequency x_1 with generations for different levels of selection, linkage and environmental variability. Figure 9 shows the exact change in the function of the gametic frequency x_1 with generations for different levels of selection, linkage and variability when $x_{1,0} = .01$ and $D_0 = 0$. In Figure 9, the function $\ln x_1$ is used to linearize the change for values of $p = x_1 + x_2 \leq .5$ and for $p > .5$ the function $\frac{1}{1-x_1} - \ln(1-x_1)$ is considered. Table 32, Appendix B, gives the exact slope β of the functions of the gametic frequency and their estimates $\hat{\beta}$ for different levels of selection, linkage and environmental variability. The values for the exact slope β are obtained from Figure 9. It is seen from the table that $\hat{\beta}$ over-estimate β when selection is strong and under-estimates β when selection is weak.

Equilibrium in the gametic frequencies exists when $\Delta x_{i,n} = 0$ for all i ($= 1, 2, 3, 4$). It is seen from (6.40) that the equilibrium condition will be satisfied by the equilibrium values $x_1 = 1$ and $x_2 = x_3 = x_4 = 0$. The approach to these equilibrium values will depend on selection and the amount of variability present.

Overdominance

Two cases will be considered, one in which only the double heterozygotes have advantage over the other genotypes which are assumed to be equal with regard to their mean phenotypic values, and in the second the heterozygote advantage is cumulative.

1. Case of double heterozygotes.

The mean phenotypic values of the genotypes are given in Table 27, Appendix B. The relation among the means of the genotypic classes may be written as

$$\mu_{22} = \mu_{21} = \mu_{12} = \mu_{20} = \mu_{02} = \mu_{10} = \mu_{01} = \mu_{00} = 0 < \mu_{11} .$$

This model accounts for interaction between the two loci, that is, it allows for epistatic effects. Accordingly, equation (6.11) takes the form

$$S = M_{11} - [1 - 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n})] \mu_{11}N_{11}(T_n) , \quad (6.45)$$

and if we use the mean phenotypic values given in Table 27, Appendix B, we find

$$S = M_{11} - 2 [1 - 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n})] N_{11}(T_n) . \quad (6.46)$$

Equation (6.18) which gives the relationship between the truncation point T_n and the gametic frequencies, reduces to

$$T_n - T_0 = 2\mu_{11} [(x_{1,n}x_{4,n} + x_{2,n}x_{3,n}) - (x_{1,0}x_{4,0} + x_{2,0}x_{3,0})] \quad (6.47)$$

For the symmetrical case, $x_1 = x_4 = \frac{1}{2} - x_2 = \frac{1}{2} - x_3$, equation (6.47) becomes

$$T_n - T_0 = 2\mu_{11} [2x_{1,n}^2 - x_{1,n} - 2x_{1,0}^2 + x_{1,0}] . \quad (6.48)$$

Equations (6.21) for the new gametic frequencies may now be written in the form

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} \{x_{1,n} [M_{11} - (1-x_{4,n})\mu_{11}N_{11}(T_n)] - rD_n M_{11}\} \\ x_{2,n+1} &= \frac{1}{S} \{x_{2,n} [M_{11} - (1-x_{3,n})\mu_{11}N_{11}(T_n)] + rD_n M_{11}\} \\ x_{3,n+1} &= \frac{1}{S} \{x_{3,n} [M_{11} - (1-x_{2,n})\mu_{11}N_{11}(T_n)] + rD_n M_{11}\} \\ x_{4,n+1} &= \frac{1}{S} \{x_{4,n} [M_{11} - (1-x_{1,n})\mu_{11}N_{11}(T_n)] - rD_n M_{11}\} . \end{aligned} \quad (6.49)$$

Equations (6.49) show the dependence of the gametic frequencies on selection, linkage and the amount of variability. When $D > 0$, the change in the frequencies of x_1 and x_4 will be increased by linkage. However, gene frequency changes independently of linkage. The frequency of A gene in the $(n+1)$ th generation is given by

$$p_{n+1} = x_{1,n+1} + x_{2,n+1} .$$

Using equations (6.49), we find

$$p_{n+1} = \frac{1}{2} - \frac{1}{2S} [M_{11} - \mu_{11} N_{11}(T_n)] (1-2p_n) . \quad (6.50)$$

Thus if $p_0 = \frac{1}{2}$, gene frequency will remain at $\frac{1}{2}$. Equation (6.50) shows that gene frequency changes independently of linkage.

The change in the gametic frequencies in one generation is given by

$$\Delta x_{i,n} = x_{i,n+1} - x_{i,n} \quad i = 1, 2, 3, 4 .$$

Using equations (6.49), we find

$$\begin{aligned}
 \Delta x_{1,n} &= \frac{1}{S} \{x_{1,n} [x_{4,n}(1-2x_{1,n}) - 2x_{2,n}x_{3,n}] \mu_{11} N_{11}(T_n) - rD_n M_{11}\} \\
 \Delta x_{2,n} &= \frac{1}{S} \{x_{2,n} [x_{3,n}(1-2x_{2,n}) - 2x_{1,n}x_{4,n}] \mu_{11} N_{11}(T_n) + rD_n M_{11}\} \\
 \Delta x_{3,n} &= \frac{1}{S} \{x_{3,n} [x_{2,n}(1-2x_{3,n}) - 2x_{1,n}x_{4,n}] \mu_{11} N_{11}(T_n) + rD_n M_{11}\} \\
 \Delta x_{4,n} &= \frac{1}{S} \{x_{4,n} [x_{1,n}(1-2x_{4,n}) - 2x_{2,n}x_{3,n}] \mu_{11} N_{11}(T_n) - rD_n M_{11}\} .
 \end{aligned} \tag{6.51}$$

The change in the gametic frequencies may be continuized by writing

$$\Delta x_{i,n} \doteq \frac{dx_{i,n}}{dn}, \quad i = 1, 2, 3, 4. \tag{6.52}$$

Because of the symmetry of equations (6.49) and (6.51), it is seen that if in the initial population $x_{1,0} = x_{4,0}$, the relation among the gametic frequencies will be $x_1 = x_4 = \frac{1}{2} - x_2 = \frac{1}{2} - x_3$ in every generation. In this case we have

$$\frac{dx_1}{dn} = \frac{dx_4}{dn} = -\frac{dx_2}{dn} = -\frac{dx_3}{dn}. \tag{6.53}$$

The gametic disequilibrium in the $(n+1)$ th generation is given by

$$\begin{aligned}
 D_{n+1} &= x_{1,n+1}x_{4,n+1} - x_{2,n+1}x_{3,n+1} \\
 &= (x_{1,n} + \frac{dx_{1,n}}{dn})(x_{4,n} + \frac{dx_{4,n}}{dn}) - (x_{2,n} + \frac{dx_{2,n}}{dn})(x_{3,n} + \frac{dx_{3,n}}{dn}).
 \end{aligned}$$

Using equations (6.52) and (6.53), we find, approximately.

$$D_{n+1} = D_n + \frac{dx_{1,n}}{dn}. \tag{6.54}$$

The change in the gametic disequilibrium is given by

$$\Delta D_n = D_{n+1} - D_n = \frac{dx_{1,n}}{dn} . \quad (6.55)$$

Equation (6.55) indicates that the conditions required for a stable equilibrium in the gametic frequencies are also required for a stable gametic equilibrium or disequilibrium. Using equations (6.51) and (6.52), we find

$$\begin{aligned} D_{n+1} &= \frac{1}{S} \{ x_{1,n} [x_{4,n}(1-2x_{1,n}) - 2x_{2,n}x_{3,n}] \mu_{11} N_{11}(T_n) - r D_n M_{11} + D_n \} \\ &= \frac{1}{S} D_n (S - r M_{11}) + \frac{1}{S} x_{1,n} \mu_{11} N_{11}(T_n) [x_{4,n} - 2(x_{1,n}x_{4,n} + x_{2,n}x_{3,n})] . \end{aligned} \quad (6.56)$$

Equation (6.56) shows that selection is likely to build up gametic disequilibrium even when $D_0 = 0$, and that linkage will increase the amount of gametic disequilibrium.

Consider the change in the frequency of x_1 given by the first equation in (6.51). Using (6.52), we find

$$\frac{dx_{1,n}}{dn} = \frac{1}{S} \{ x_{1,n} [x_{4,n}(1-2x_{1,n}) - 2x_{2,n}x_{3,n}] \mu_{11} N_{11}(T_n) - r D_n M_{11} \} .$$

Using equations (6.28) and (6.53), we find

$$\frac{dx_{1,n}}{dn} = \frac{1}{S} \left\{ \left[2x_{1,n} - x_{1,n}^2 - 2x_{1,n}^3 - \frac{1}{2} \right] \frac{\mu_{11}}{\sigma\sqrt{2\pi}} - r D_n M_{11} \right\} . \quad (6.57)$$

If either D or r is zero, equation (6.57) yields the following solution

$$\int_{x_{1,0}}^{x_{1,n}} \frac{1}{-\frac{1}{2} + 2x_1 - x_1^2 - 2x_1^3} dx_1 = \frac{\mu_{11}}{S\sigma\sqrt{2\pi}} n . \quad (6.58)$$

The function of the gametic frequency x_1 resulting from the solution given by (6.58) is expected to change linearly with generations with slope given by

$$\hat{\beta} = \frac{1}{S\sigma \sqrt{2\pi}} . \quad (6.59)$$

The slope $\hat{\beta}$ increases when selection is effective; that is, when both S and σ are small.

Equilibrium in the gametic frequencies exists when $\Delta x_{i,n} = 0$ for all i ($= 1, 2, 3, 4$). If either D or r is zero, it is seen from equations (6.51) that the equilibrium condition is satisfied for $x_1 = 0$, $\frac{1}{4}$ and $\frac{1}{2}$. For the equilibrium value $x_1 = \frac{1}{4}$, the gametic disequilibrium is zero. If $D \neq 0$, the equilibrium points are those which satisfy the cubic equation

$$\begin{aligned} \frac{1}{4} r M_{11} - \left(\frac{1}{2} \mu_{11} N_{11}(T_n) + r M_{11} \right) x_1 \\ + 3\mu_{11} N_{11}(T_n) x_1^2 - 4\mu_{11} N_{11}(T_n) x_1^3 = 0 . \end{aligned} \quad (6.60)$$

It is seen from (6.60) that, when $D \neq 0$, the equilibrium points will be a function of linkage and the amount of variability. Lewontin and Kojima (1960) examined the effect of natural selection on the gametic frequencies using constant symmetric selective values for the genotypes with no variability. The selective values of the genotypes are presented in Table 26.

Table 26. Selective values of the genotypes for the symmetrical case (Lewontin and Kojima, 1960).

	AA	Aa	aa
BB	a	b	a
Bb	c	d	c
bb	a	b	a

Their model corresponds to the present model if we put $d > a = b = c = 0$. In their study, the equilibrium points are those which satisfy the cubic equation

$$x_1(b + c - d - a)(4x_1^2 - 3x_1 + \frac{1}{2}) - rd(x_1 - \frac{1}{4}) = 0 \quad (6.61)$$

A comparison between (6.60) and (6.61) shows the dependence of the equilibrium and the approach to it on linkage and the amount of variability present when selective values are frequency dependent.

2. Case of cumulative heterozygote advantage.

The mean phenotypic values are given in Table 27, Appendix B. With μ_{00} coded as 0 and μ_{11} coded as 2, the values of μ_{22} , μ_{02} , and μ_{20} are 0, and of μ_{10} , μ_{01} , μ_{12} and μ_{21} are unity. This model does not involve interaction between loci. Accordingly, equation (6.11) takes the form

$$S = M_{11} - x_{1,n}(1-x_{4,n}) + x_{4,n}(1-x_{1,n}) + (x_{2,n}^2 + x_{3,n}^2)\mu_{11}N_{11}(T_n). \quad (6.62)$$

Equation (6.18) which gives the relationship between the truncation point T_n and the gametic frequencies, reduces to

$$T_n - T_0 = \mu_{11} \{ [x_{1,0}(1-x_{4,0}) + x_{4,0}(1-x_{1,0}) + (x_{2,0}^2 + x_{3,0}^2)] - [x_{1,n}(1-x_{4,n}) + x_{4,n}(1-x_{1,n}) + (x_{2,n}^2 + x_{3,n}^2)] \} . \quad (6.63)$$

Equations (6.21) for the new gametic frequencies may now be written in the form

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} \{ x_{1,n} [M_{11} - \frac{1}{2} (1 + x_{1,n} - x_{4,n}) \mu_{11} N_{11}(T_n)] - r D_n M_{11} \} \\ x_{2,n+1} &= \frac{1}{S} \{ x_{2,n} [M_{11} - \frac{1}{2} (1 + x_{2,n} - x_{3,n}) \mu_{11} N_{11}(T_n)] + r D_n M_{11} \} \\ x_{3,n+1} &= \frac{1}{S} \{ x_{3,n} [M_{11} - \frac{1}{2} (1 + x_{3,n} - x_{2,n}) \mu_{11} N_{11}(T_n)] + r D_n M_{11} \} \\ x_{4,n+1} &= \frac{1}{S} \{ x_{4,n} [M_{11} - \frac{1}{2} (1 + x_{4,n} - x_{1,n}) \mu_{11} N_{11}(T_n)] - r D_n M_{11} \} . \end{aligned} \quad (6.64)$$

The symmetry of equations (6.64) requires that $x_1 = x_4 = \frac{1}{2} - x_2 = \frac{1}{2} - x_3$, and for this case, equations (6.64) reduce to

$$\begin{aligned} x_{1,n+1} &= \frac{1}{S} \{ x_{1,n} [M_{11} - \frac{1}{2} \mu_{11} N_{11}(T_n)] - r D_n M_{11} \} \\ x_{2,n+1} &= \frac{1}{S} \{ x_{2,n} [M_{11} - \frac{1}{2} \mu_{11} N_{11}(T_n)] + r D_n M_{11} \} \\ x_{3,n+1} &= \frac{1}{S} \{ x_{3,n} [M_{11} - \frac{1}{2} \mu_{11} N_{11}(T_n)] + r D_n M_{11} \} \\ x_{4,n+1} &= \frac{1}{S} \{ x_{4,n} [M_{11} - \frac{1}{2} \mu_{11} N_{11}(T_n)] - r D_n M_{11} \} . \end{aligned} \quad (6.65)$$

The new gene frequency of A is given by

$$p_{n+1} = x_{1,n+1} + x_{2,n+1} .$$

Using equations (6.65), we find

$$p_{n+1} = \frac{1}{S} p_n \left[M_{11} - \frac{1}{2} \mu_{11} N_{11}(T_n) \right] . \quad (6.66)$$

Equation (6.66) shows that gene frequency changes independently of linkage. The change in gene frequency is given by

$$\Delta p_n = p_{n+1} - p_n .$$

Using equation (6.66), we find

$$\Delta p_n = \frac{1}{S} p_n \left[M_{11} - \frac{1}{2} \mu_{11} N_{11}(T_n) - S \right] . \quad (6.67)$$

For the symmetrical case, $x_1 = x_4 = \frac{1}{2} - x_2 = \frac{1}{2} - x_3$, equation (6.62) reduces to

$$S = M_{11} - \frac{1}{2} \mu_{11} N_{11}(T_n) . \quad (6.68)$$

Using (6.68), it is seen from (6.67) that $\Delta p_n = 0$, indicating that gene frequency will not change from its initial value of .5 under the symmetrical case.

The change in the gametic frequency in one generation is given by

$$\Delta x_{i,n} = x_{i,n+1} - x_{i,n} .$$

Using equations (6.64), we find

$$\begin{aligned} \Delta x_{1,n} = \frac{1}{S} \{ & \frac{1}{2} x_{1,n} [2x_{4,n}(1-2x_{1,n}) - (1 - x_{1,n} - x_{4,n}) \\ & + 2(x_{2,n}^2 + x_{3,n}^2)] \mu_{11} N_{11}(T_n) - r D_n M_{11} \} \end{aligned}$$

$$\begin{aligned}
\Delta x_{2,n} &= \frac{1}{S} \left\{ \frac{1}{2} x_{2,n} \left[2x_{4,n}(1 - x_{1,n} - x_{4,n}) - (1-2x_{1,n}) - (x_{2,n} - x_{3,n}) \right. \right. \\
&\quad \left. \left. + 2(x_{2,n}^2 + x_{3,n}^2) \right] \mu_{11} N_{11}(T_n) + r D_n M_{11} \right\} \\
\Delta x_{3,n} &= \frac{1}{S} \left\{ \frac{1}{2} x_{3,n} \left[2x_{4,n}(1 - x_{1,n} - x_{4,n}) - (1-2x_{1,n}) - (x_{3,n} - x_{2,n}) \right. \right. \\
&\quad \left. \left. + 2(x_{2,n}^2 + x_{3,n}^2) \right] \mu_{11} N_{11}(T_n) + r D_n M_{11} \right\} \quad (6.69) \\
\Delta x_{4,n} &= \frac{1}{S} \left\{ \frac{1}{2} x_{4,n} \left[2(x_{1,n} - x_{4,n}) - (1-2x_{4,n})(1 - x_{1,n} - x_{4,n}) \right. \right. \\
&\quad \left. \left. + 2(x_{2,n}^2 + x_{3,n}^2) \right] \mu_{11} N_{11}(T_n) - r D_n M_{11} \right\} .
\end{aligned}$$

For the symmetrical case and using equations (6.65), we get

$$\begin{aligned}
\Delta x_{1,n} &= \frac{1}{S} \left(\frac{1}{4} - x_{1,n} \right) r M_{11} \\
\Delta x_{2,n} &= \frac{1}{S} \left(x_{1,n} - \frac{1}{4} \right) r M_{11} \\
\Delta x_{3,n} &= \frac{1}{S} \left(x_{1,n} - \frac{1}{4} \right) r M_{11} \\
\Delta x_{4,n} &= \frac{1}{S} \left(\frac{1}{4} - x_{1,n} \right) r M_{11} .
\end{aligned} \quad (6.70)$$

Equilibrium exists when $\Delta x_{i,n} = 0$ for all i . For the symmetrical case, it is seen from (6.70) that there is only one equilibrium given by $x_1 = x_2 = x_3 = x_4 = \frac{1}{4}$ and it is stable.

The change in the gametic frequencies may be continuized by writing

$$\Delta x_{i,n} \doteq \frac{dx_{i,n}}{dn} \quad i = 1, 2, 3, 4 .$$

Using equations (6.70) for the symmetrical case, we find

$$\frac{dx_{1,n}}{dn} = \frac{1}{S} \left(\frac{1}{4} - x_{1,n} \right) r M_{11} . \quad (6.71)$$

For the symmetrical case, equation (6.63) becomes

$$T_n - T_0 = 0 ,$$

and M_{11} defined in (6.4) becomes

$$\begin{aligned} M_{11}(T_n) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{T_n}^{\infty} \exp \left[-\frac{(t-\mu_{11})^2}{2\sigma^2} \right] dt \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{T_0}^{\infty} \exp \left[-\frac{(t-\mu_{11})^2}{2\sigma^2} \right] dt \\ &= M_{11}(T_0) \end{aligned} \quad (6.72)$$

Equation (6.72) shows that M_{11} is constant. Using (6.72), equation (6.71) yields the following solution

$$- \ln\left(\frac{1}{4} - x_{1,n}\right) = - \ln\left(\frac{1}{4} - x_{1,0}\right) + \frac{rM_{11}(T_0)}{S} n . \quad (6.73)$$

Case 2: σ is very small.

In this case we may write equation (6.16) in the approximate form

$$\frac{(T_n - \mu_{11})}{\sigma^2} f_n(\underline{x}, \underline{\mu}) dT_n = [f'_n(\underline{x}, \underline{\mu}) d\bar{x}] , \quad (6.74)$$

where $f_n(\underline{x}, \underline{\mu})$ and $[f'_n(\underline{x}, \underline{\mu}) d\bar{x}]$ are defined in (6.12) and (6.15), respectively.

Equation (6.74) may be written in the form

$$\frac{(T_n - \mu_{11})^2}{\sigma^2} dT_n = \frac{f'_n(\underline{x}, \underline{\mu})}{f_n(\underline{x}, \underline{\mu})} d\underline{x} . \quad (6.75)$$

Equation (6.75) yields the following solution

$$\frac{(T_n - \mu_{11})^2}{2\sigma^2} - \frac{(T_0 - \mu_{11})^2}{2\sigma^2} = \ln f_n(\underline{x}, \underline{\mu}) - \ln f_0(\underline{x}, \underline{\mu}) , \quad (6.76)$$

where $f_n(\underline{x}, \underline{\mu})$ is defined in (6.12). Using (6.76), the term $N_{11}(T_n)$ defined in (6.3) becomes

$$\begin{aligned} N_{11}(T_n) &= \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(T_n - \mu_{11})^2}{2\sigma^2} \right] \\ &= \frac{1}{\sigma\sqrt{2\pi}} \frac{f_0(\underline{x}, \underline{\mu})}{f_n(\underline{x}, \underline{\mu})} \exp \left[-\frac{(T_0 - \mu_{11})^2}{2\sigma^2} \right] . \end{aligned} \quad (6.77)$$

For the case of no dominance of both loci, the change in the frequency x_1 given by the first equation in (6.25) may be approximated by

$$\frac{dx_{1,n}}{dn} = \frac{1}{S} \left\{ \frac{1}{2} x_{1,n} (1 - x_{1,n} + x_{4,n}) \mu_{11} N_{11}(T_n) - r D_n M_{11} \right\} .$$

Using equation (6.77), we find

$$\begin{aligned} \frac{dx_{1,n}}{dn} &= \frac{\mu_{11}(x_{1,0} - x_{4,0})}{2S\sigma\sqrt{2\pi}} \exp \left[-\frac{(T_0 - \mu_{11})^2}{2\sigma^2} \right] \\ &\quad \frac{x_{1,n}(1 - x_{1,n} + x_{4,n})}{x_{1,n} - x_{4,n}} - r D_n M_{11}/S . \end{aligned} \quad (6.78)$$

If either D or r is zero and x_4 is small, equation (6.78) may be written in the approximate form

$$\frac{dx_{1,n}}{dn} = \frac{\mu_{11}(x_{1,0} - x_{4,0})}{2S \sigma \sqrt{2\pi}} \exp \left[-\frac{(T_0 - \mu_{11})^2}{2\sigma^2} \right] (1 - x_{1,n}) . \quad (6.79)$$

Equation (6.79) yields the following solution

$$\ln \frac{1}{1 - x_{1,n}} = \ln \frac{1}{1 - x_{1,0}} + \frac{\mu_{11}(x_{1,0} - x_{4,0}) \exp \left[-(T_0 - \mu_{11})^2 / 2\sigma^2 \right]}{2S \sigma \sqrt{2\pi}} n . \quad (6.80)$$

The function of the gametic frequency x_1 given by the left-hand side of (6.80) is expected to change linearly with generations with slope given by

$$\hat{\beta} = \frac{\mu_{11}(x_{1,0} - x_{4,0}) \exp \left[-(T_0 - \mu_{11})^2 / 2\sigma^2 \right]}{2S \sigma \sqrt{2\pi}} . \quad (6.81)$$

Solutions similar to (6.80) can be obtained for other models by evaluating $f_n(\underline{x}, \underline{\mu})$ defined in (6.12) under the given model, substituting for $N_{11}(T_n)$ in the expression of the change in the gametic frequency, and then solving the resultant differential equation. However, such a solution is difficult to obtain in most of the cases without making further simplifications.

EXTENSION TO n LOCI

The present approach has the advantage of simplifying the notation to a considerable extent, while nevertheless retaining enough generality in many circumstances to allow straightforward generalization to the case of n loci with arbitrary number of alleles. However, the work is expected to involve a considerable amount of mathematics.

For the case of more than two loci with two or more alleles at each locus, the number of distinct possible phenotypes as well as the dominance relations among the means of the genotypic classes will be very large. Hence the number of possible stages for selection will be increased tremendously. For example, with three loci segregating and two alleles at each locus, A, a, B, b, C, c, there are eight possible gametic types ABC, ABc, AbC, aBC, Abc, aBc, abC and abc with frequencies $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$, respectively. If the coupling and repulsion double as well as triple heterozygotes are assumed to be equally viable, there are 27 distinct possible phenotypic classes. Accordingly, for a given dominance relation, there will be 27 possible stages of progress. If we do not assume independence of the two intervals A — B and B — C; that is, if we allow interference, there are $2^{3-1} = 4$ linkage parameters made up of the non-crossovers, and single and double exchanges. These linkage parameters (probabilities) add to unity however, so there are only three independent linkage parameters consisting of two single and one double exchange.

The measure of gametic disequilibrium D can be extended to the case of more than two loci by computing separate D values for each pair of loci. Thus the gametic disequilibrium for the three possible pairs of loci in the three locus case would be

$$D_{12} = (x_1 + x_2)(x_8 + x_7) - (x_3 + x_5)(x_4 + x_6)$$

$$D_{13} = (x_1 + x_3)(x_8 + x_6) - (x_2 + x_5)(x_4 + x_7)$$

$$D_{23} = (x_1 + x_4)(x_8 + x_5) - (x_2 + x_6)(x_3 + x_7) .$$

For the case of environmental variability, the foregoing approach can, perhaps, be extended to the case of n loci. While the development of a theory for an arbitrary number of loci which will yield a curve of response to continued selection is greatly needed, the difficulties are considerable.

SUMMARY AND CONCLUSIONS

The theory of the progress of infinite populations under the culling or truncation type of selection, in contrast to the study only of equilibria conditions, was examined in detail for the case of one locus with two and three alleles and the case of two loci with two alleles at each locus. In the theory which has been discussed in the present work, the fraction of the population saved was the best S ($0 < S < 1$) with regard to the attribute under selection. The basic idea in the development is that there is a scale of merit and all above a truncation point are selected and all below this point are culled. With this model of selection, the actual dependence of the 'selective values' of genotypes is determined by the relative positions of the genotypes on a scale of merit and by the relative frequencies. The present work can therefore be regarded as a study of gene frequency dependent selection. The progress of the population under this model depended on the order of merit of the genotypes.

The main difficulty in the development arises because the degree of selection against any one genotype depends critically on the population structure at the time of selection. The purpose of the present study was to examine the dynamics of a genetic population under the culling type of selection, both without and with environmental variability in the expression on the scale of merit of genotypes.

In order to keep the mathematics workable, two life phases---infant and adult, non-overlapping generations, no differences between the two sexes in fitness, random mating and infinite population size were assumed. It was also assumed that there are no fecundity differences between the

possible genetic types of mating. The validity of this type of simplifying approach derives from a feeling that in the advance of science it is not simplification that leads to error, but rather the absence of a rigorous and clear analysis of the problem at hand. A simplified approach, however, may also reveal whether a given theory holds enough promise to warrant further investigation.

It is seen, even in some simple cases, the case of one locus with three alleles and the two loci with two alleles at each locus, that there are many possible genotypic or phenotypic configurations with many possibilities of dominance relations of the genotypic values; each leads to a different overall response relation of the population to the number of generations. The present work was therefore confined to two cases, one in which heterozygotes were assumed to be superior over the homozygotes and the other in which intermediate and complete dominance hold.

For each model considered, the selection experiment was programmed on the computer and the progress of the population was evaluated with no approximations used. Thus, the computer results were considered to be exact and were, therefore, used to examine the accuracy of the foregoing mathematical equations.

For the case of no variability, the phenotypes associated with the genotypes were assumed to be constant with no variation due to non-genetic causes. The progress of the population under this model depended on the order of merit of the genotypes. The progress under selection proceeded in stages, with the number of stages dependent on the number of distinct possible phenotypes and on the proportion culled in relation to gene frequency. The ordering of the phenotypic classes was assumed to be the

same in the two sexes. For the two-locus model, the coupling and repulsion double heterozygotes were assumed to be equally viable. Linkage between the two loci was assumed to be the same for the two sexes. This simplification often does not exist. For example, an extreme case occurs in *Drosophila* where crossing over does not occur in the male. Thus, a more complicated analysis is required to accommodate different recombination values in the two sexes. However, such a complication does not change the general picture; it merely alters the speed of the response to selection.

In general, it was found that, in the absence of overdominance, the mean genotype of the population is non-decreasing from one generation to the next and it reaches its maximum at the equilibrium frequencies. In this case, equilibrium results in fixation of the most desirable gene. When overdominance is present, a stable equilibrium between 0 and 1 can be attained under some cases of selection pressure. For a given situation, the type of equilibrium depended on the relative proportions culled. For the overdominance model and from the results of the one locus with two and three alleles, it was found that stable equilibrium frequencies do not necessarily correspond to maximum mean genotypic value of the population. These results differ from those that would be expected from the classical theory discussed by Fisher, in which a random mating population has maximum mean fitness when the gene frequencies are at stable equilibrium values.

In some simple cases in which exact solutions were obtained, a comparison was made between the change in gene frequency as a discrete process expressed in terms of difference equations and the change obtained using the continuous time analog. It was found that the approximation using the

continuous model is only accurate when selection is very weak ($S > .99$). When selection was strong, large differences were found between the two approaches.

In most cases, functions of gene frequencies were obtained which change linearly with generations. The slopes of these linear relationships were dependent on the initial structure of the population and the parameters involved.

In the two loci model without variability, it was found that linkage may have profound effect on the change in the gametic frequencies. The effect of linkage depends on the magnitude of the gametic disequilibrium. With intermediate and complete dominance models, it was found that selection may build up negative gametic disequilibrium even when the initial population is in gametic equilibrium and there is no linkage. Thus, for those models the effect of linkage will be to accelerate the rate of change in the frequency of the most desirable gametes when linkage is tight and retard the rate when linkage is loose. For the case in which the two loci were assumed to have equal effects, it was found that the frequencies of the two genes will remain equal in every generation if their initial frequencies were the same, and that the difference between the frequencies of the two genes will be increased by selection if the frequencies differ in the initial population. For the case in which the two loci were assumed to have unequal effects, it was found that selection will increase the difference between the frequencies of the two genes even if the frequencies were the same in the initial population. For the pure overdominance model with no interaction between the two loci, it was found that stable equilibrium results in equal frequencies of the four gametes. While it was found

that linkage may have some effects on the changes in the gametic frequencies, gene frequencies change independently of linkage. With overdominance and in the presence of epistatic effects, a stable equilibrium with equal frequencies of the four gametes can be attained when either the recombination frequency is greater than .25 or there is no gametic disequilibrium. In the presence of gametic disequilibrium, a stable equilibrium can be attained if linkage is tighter than .25.

The cases of one and two loci with environmental variability were also examined. In these cases, the expressions of the genotypes were assumed to vary so that the value on the scale of merit of an individual with a given genotype was normally distributed around some fixed value μ with variance σ^2 . For simplicity, the variances associated with the genotypes were assumed to be the same for all genotypes. The progress of the population with given levels of the parameters considered was examined as well as the relationships between the effects of these parameters on the equilibrium that is attained and the approach to the equilibrium. For a given situation, the progress of the population was expressed as a function of the degree of selection, the amount of variability present, linkage, differences between the means of the phenotypic classes, and the initial structure of the population. Two approximations were considered and compared with other approximate solutions and exact solutions obtained by a computer.

BIBLIOGRAPHY

- Bodmer, W. F. 1965. Differential fertility in population genetic models. *Genetics* 51: 411-424.
- Bodmer, W. F. and J. Felsenstein. 1967. Linkage and selection: theoretical analysis of the deterministic two locus random mating model. *Genetics* 57: 237-265.
- Clark, B. and P. O'Donald. 1964. Frequency dependent selection. *Heredity* 19: 201-206.
- Crow, J. F. and M. Kimura. 1969. *An Introduction to Population Genetics Theory*. Harper and Row Publishers, New York, New York.
- Curnow, R. N. 1964. The effect of continued selection of phenotypic intermediates on gene frequency. *Genetic Research* 5: 341-353.
- Deakin, M. A. B. 1968. Interaction of selection and linkage. *Australian Journal of Biological Sciences* 21: 55-59.
- Dobzhansky, T. and B. Spassky. 1967. An experiment on migration and simultaneous selection for several traits in *Drosophila pseudoobscura*. *Genetics* 55: 723-734.
- Ewens, W. J. 1969a. A generalized fundamental theory of natural selection. *Genetics* 63: 531-539.
- Ewens, W. J. 1969b. *Population Genetics*. Barnes and Noble Incorporated, New York, New York.
- Falconer, D. S. 1957. Selection for phenotypic intermediates in *Drosophila*. *Journal of Genetics* 55: 551-561.
- Falconer, D. S. and A. Robertson. 1956. Selection for environmental variability of body size in mice. *A. Indukt. Abstamm.-u. VererbLehre* 87: 385-391.
- Felsenstein, J. 1965. The effect of linkage on directional selection. *Genetics* 52: 349-363.
- Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions Royal Society Edinburgh* 52: 399-433.
- Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford, England.

- Griffing, B. 1960. Theoretical consequences of truncation selection based on the individual phenotype. *Australian Journal of Biological Sciences* 13: 307-343.
- Haldane, J. B. S. 1924. The mathematical theory of natural and artificial selection. Part I. *Transactions Cambridge Philosophical Society* 23: 19-41.
- Haldane, J. B. S. 1926. The mathematical theory of natural and artificial selection. Part III. *Proceedings Cambridge Philosophical Society* 23: 363-372.
- Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection. Part IV. *Proceedings Cambridge Philosophical Society* 23: 607-615.
- Haldane, J. B. S. 1931. A mathematical theory of natural and artificial selection. Part VII. Selection intensity as a function of mortality rate. *Proceedings Cambridge Philosophical Society* 27: 131-136.
- Haldane, J. B. S. 1932. *The Causes of Evolution*. Longmans, Green and Company, London.
- Haldane, J. B. S. 1961. Some simple systems of artificial selection. *Journal of Genetics* 56: 345-350.
- Haldane, J. B. S. 1962. The selection of double heterozygotes. *Journal of Genetics* 58: 125-128.
- Karlin, S. and F. M. Scudo. 1969. Assortative mating based on phenotype. II. Two autosomal alleles without dominance. *Genetics* 63: 499-510.
- Kempthorne, O. 1957. *An Introduction to Genetic Statistics*. John Wiley and Sons, Incorporated, New York, New York.
- Kempthorne, O. and E. Pollak. 1970. Concepts of fitness in Mendelian populations. *Iowa Agriculture and Home Economics Experiment Station Journal Paper J-1669*.
- Kimura, M. 1956. A model of a genetic system which leads to closer linkage by natural selection. *Evolution* 10: 278-287.
- Kimura, M. 1958. On the change of population fitness by natural selection. *Heredity* 12: 145-167.
- Kimura, M. 1965. Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. *Genetics* 52: 875-890.
- Kojima, K. and T. M. Kelleher. 1960. Changes of mean fitness in random mating populations when epistasis and linkage are present. *Genetics* 46: 527-540.

- Kojima, K. K. and K. Yarbrough. 1967. The mode of selection at the polymorphic *Estrase 6* locus in cage populations of *Drosophila melanogaster*. *Genetics* 57: 677-686.
- Latter, B. D. H. 1965. The response to artificial selection due to autosomal genes of large effect. I. Changes in gene frequency at an additive locus. *Australian Journal of Biological Sciences* 18: 585-598.
- Lewontin, R. C. 1958. A general method for investigating the equilibrium of gene frequency in a population. *Genetics* 43: 419-434.
- Lewontin, R. C. 1964a. The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* 49: 49-67.
- Lewontin, R. C. 1964b. The interaction of selection and linkage. II. Optimum models. *Genetics* 50: 757-782.
- Lewontin, R. C. and Ken-Ichi Kojima. 1960. The evolutionary dynamics of complex polymorphisms. *Evolution* 14: 458-472.
- Li, C. C. 1955. *Population Genetics*. The University of Chicago Press, Chicago, Illinois.
- Li, C. C. 1967. Genetic equilibrium under selection. *Biometrics* 23: 397-484.
- Mandel, S. P. H. 1958. The stability of a multiple allelic system. *Heredity* 12: 289-302.
- Moran, P. A. P. 1962. *The Statistical Processes of Evolutionary Theory*. Clarendon Press, Oxford, England.
- Moran, P. A. P. 1963. On the measurement of natural selection dependent on several loci. *Evolution* 17: 182-186.
- Moran, P. A. P. 1967. Unsolved problems in evolutionary theory. *Proceedings of the Fifth Berkley Symposium on Mathematical Statistics and Probability* 4: 457-480.
- O'Donald, P. 1960. Assortive mating in a population in which two alleles are segregating. *Heredity* 15: 389-396.
- Parsons, P. A. 1963. Polymorphism and the balanced polygenic combination. *Evolution* 17: 564-574.
- Penrose, L. S. 1949. The meaning of fitness in human populations. *Annals of Eugenics*, London, 14: 301-304.
- Pollak, E. 1966. Some consequences of selection by culling when there is superiority of heterozygotes. *Genetics* 53: 977-988.

- Prout, T. 1962. The effects of stabilizing selection in the time of development in *Drosophila melanogaster*. *Genetical Research* 3: 364-382.
- Robertson, A. 1956. The effect of selection against extreme deviants based on deviation or on homozygosis. *Journal of Genetics* 54: 236-248.
- Robertson, A. 1960. A theory of limits in artificial selection. *Proceedings of the Royal Society Series B*. 153: 234-249.
- Tallis, G. M. 1966. Equilibria under selection for K alleles. *Biometrics* 22: 121-127.
- Thomson, J. A. 1961. Interallelic selection in experimental populations of *Drosophila melanogaster*: white and satsuma. *Genetics* 46: 1435-1442.
- Van Der Veen, J. H. 1960. Heterozygote superiority, selection intensity and plateauing. *Heredity* 15: 321-323.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97-159.
- Wright, S. 1942. Statistical genetics and evolution. *Bulletin of the American Mathematical Society* 48: 223-246.
- Wright, S. 1969. Evolution and the Genetics of Populations. A Treatise in three volumes. Volume II. The Theory of Gene Frequencies. The University of Chicago Press, Chicago, Illinois.

ACKNOWLEDGEMENTS

The author wishes to express his deep and sincere appreciation to Professor Oscar Kempthorne for the opportunity of being apprenticed to him in learning. The author wishes to thank Professor P. A. Peterson for teaching, stimulation and encouragement in the field of genetics. Thanks also to Dr. E. Pollak for his reviews and discussions.

The author wishes to thank Mrs. Marlene Sposito for her skill in preparing this manuscript.

The work was part of a research effort which was supported by the National Institute of Health under Grant Number 430-43-53-81-1669.

APPENDIX A: FIGURES

Figure 1. The exact change in gene frequency (p) with generations (n) under different levels of selection (S). Initial frequency $p_0 = .05$. One Locus-No Variability.

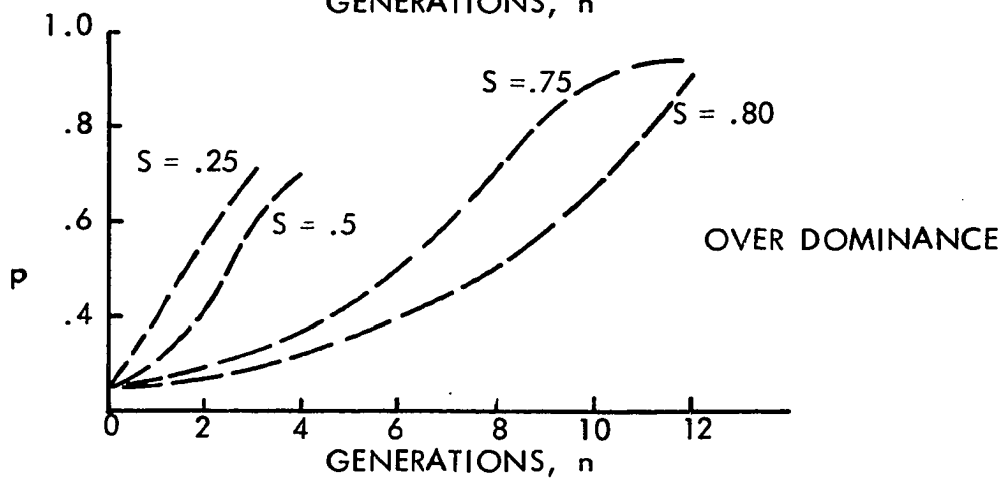
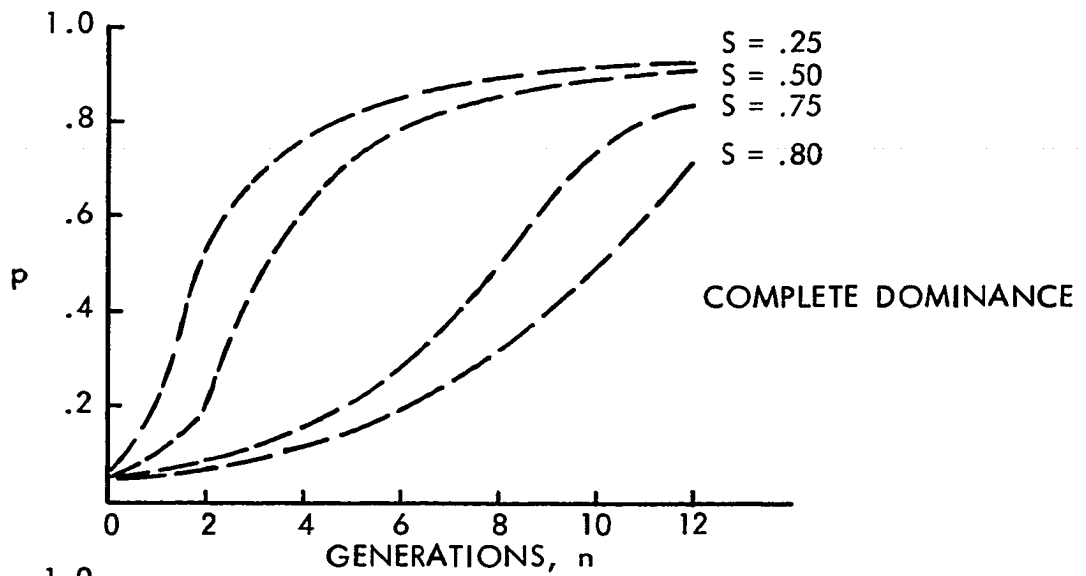
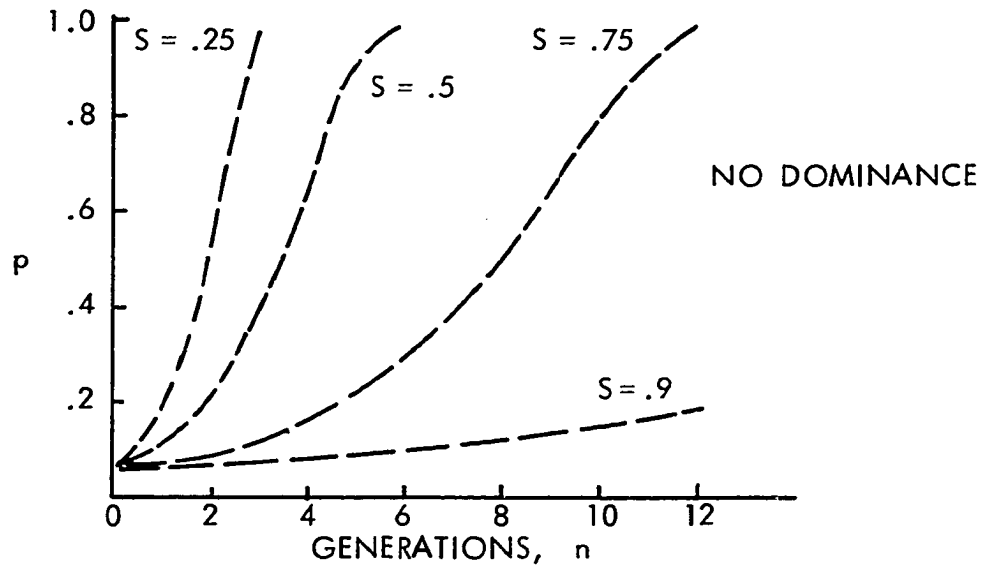


Figure 2. The exact change in the function of gene frequency (p) with generations (n) under different levels of selection (S).

Initial frequency $p_0 = .05$. One Locus-No Variability.

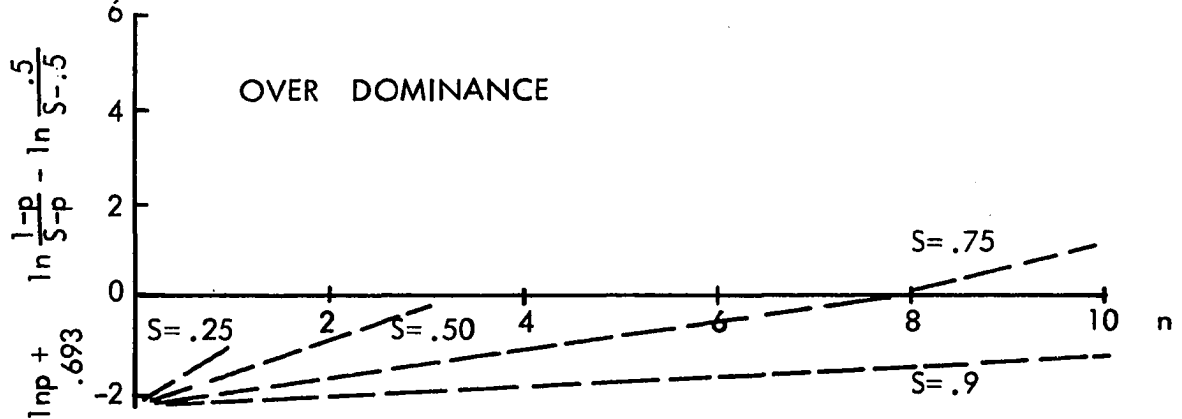
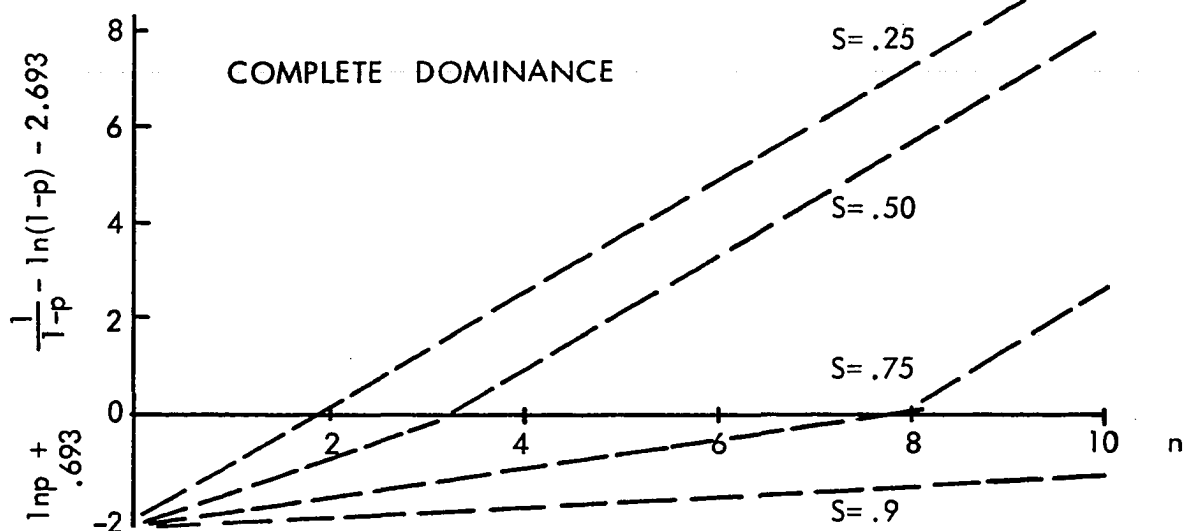
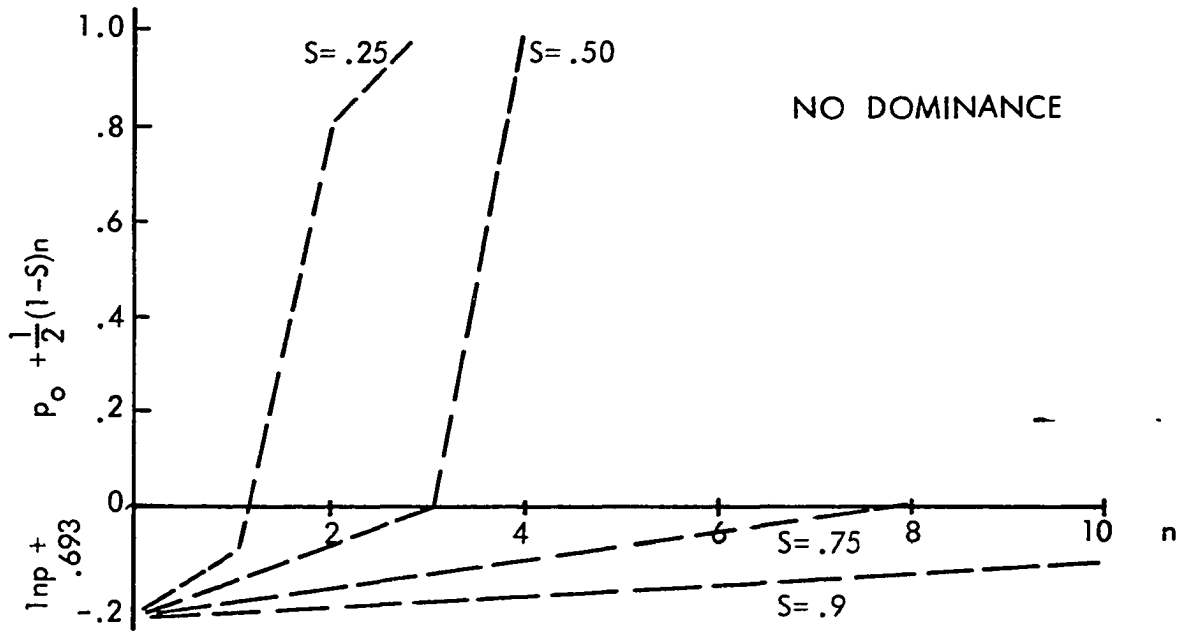


Figure 3. The exact change in gametic disequilibrium (D) with generations (n) under different levels of selection (S) and linkage (r). Initial frequency $p_0 = q_0 = .1$ and $D_0 = 0$. Two Loci-
No Variability-No Dominance.

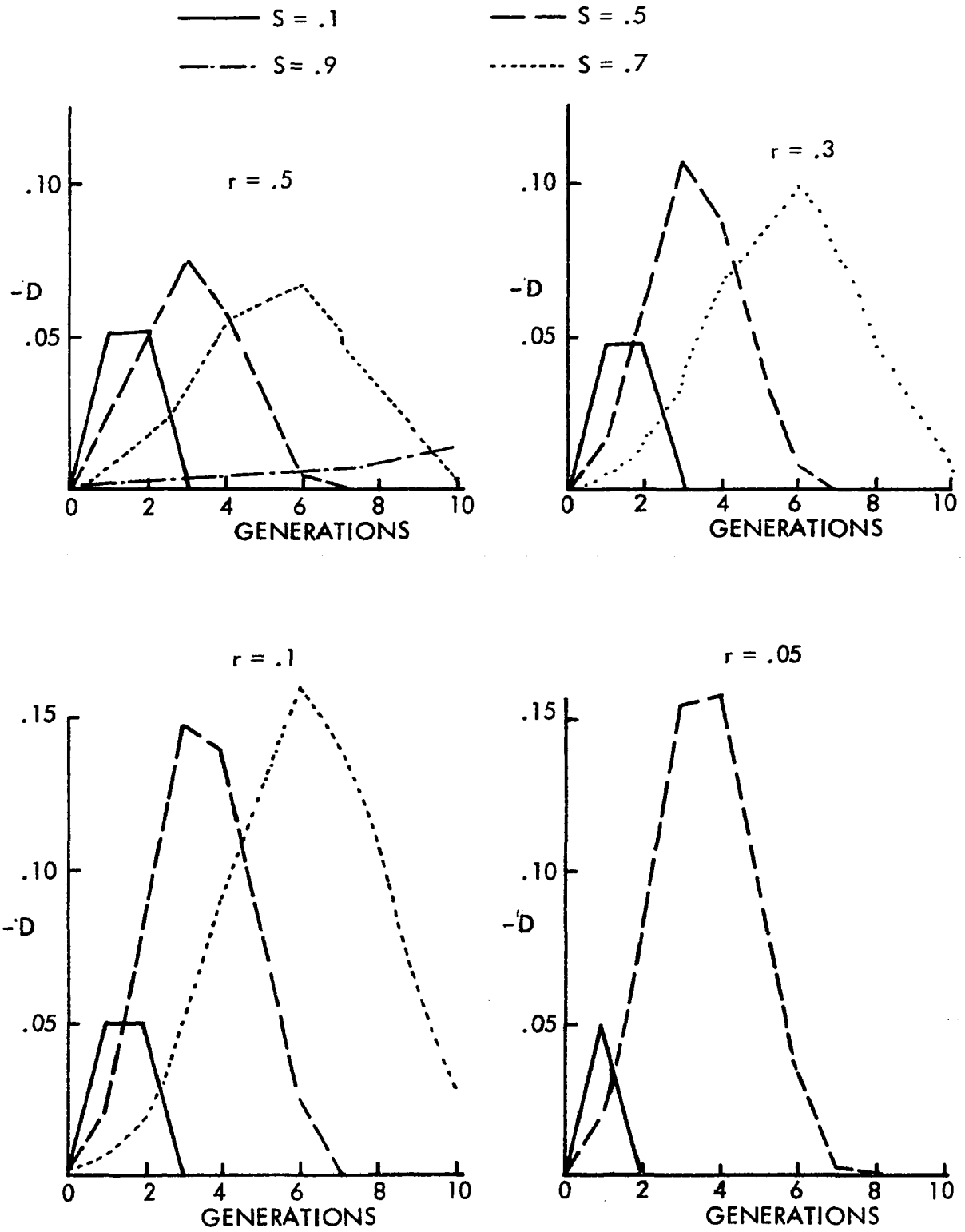


Figure 4. The exact change in the function of the gametic frequency x_1 with generations (n) under different levels of selection (S) and linkage (r). Initial frequency $x_{1,0} = .01$ and $D_0 = 0$. Two Loci-No Variability-No Dominance.

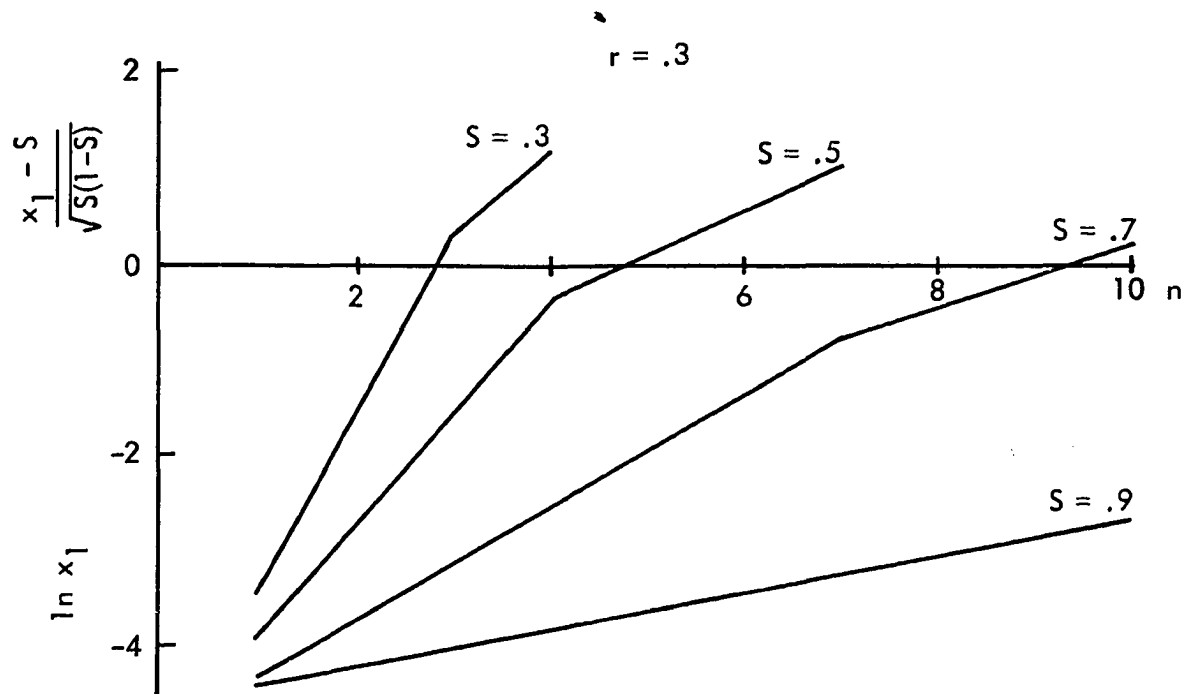
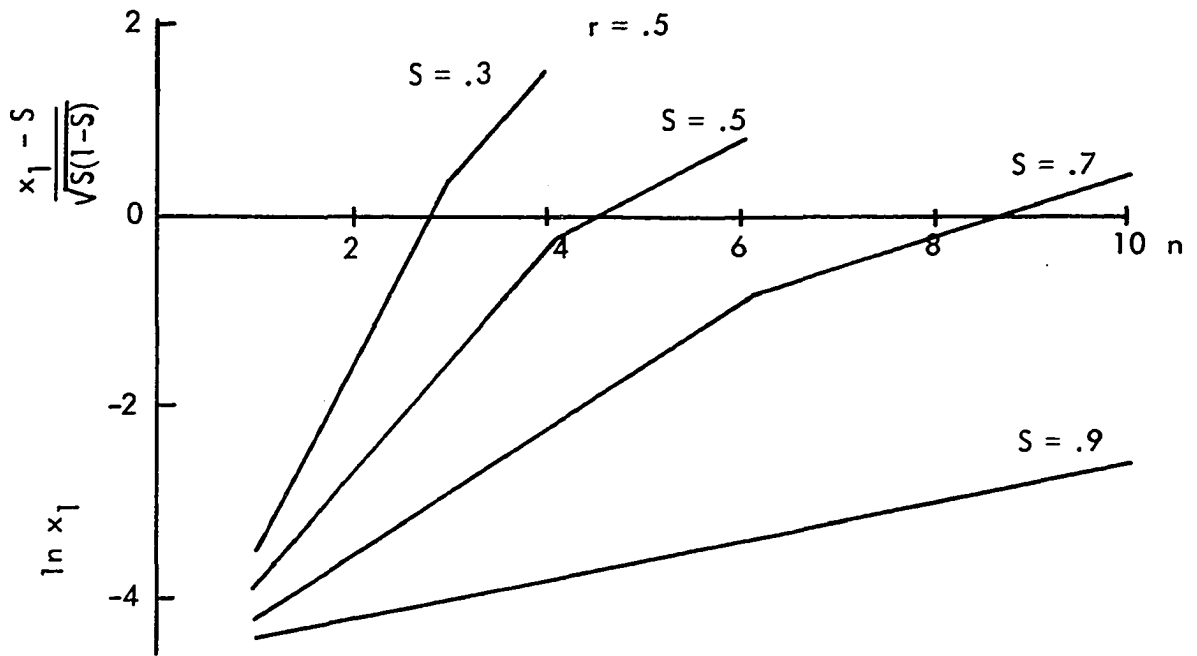


Figure 4 (continued) - Complete Dominance.

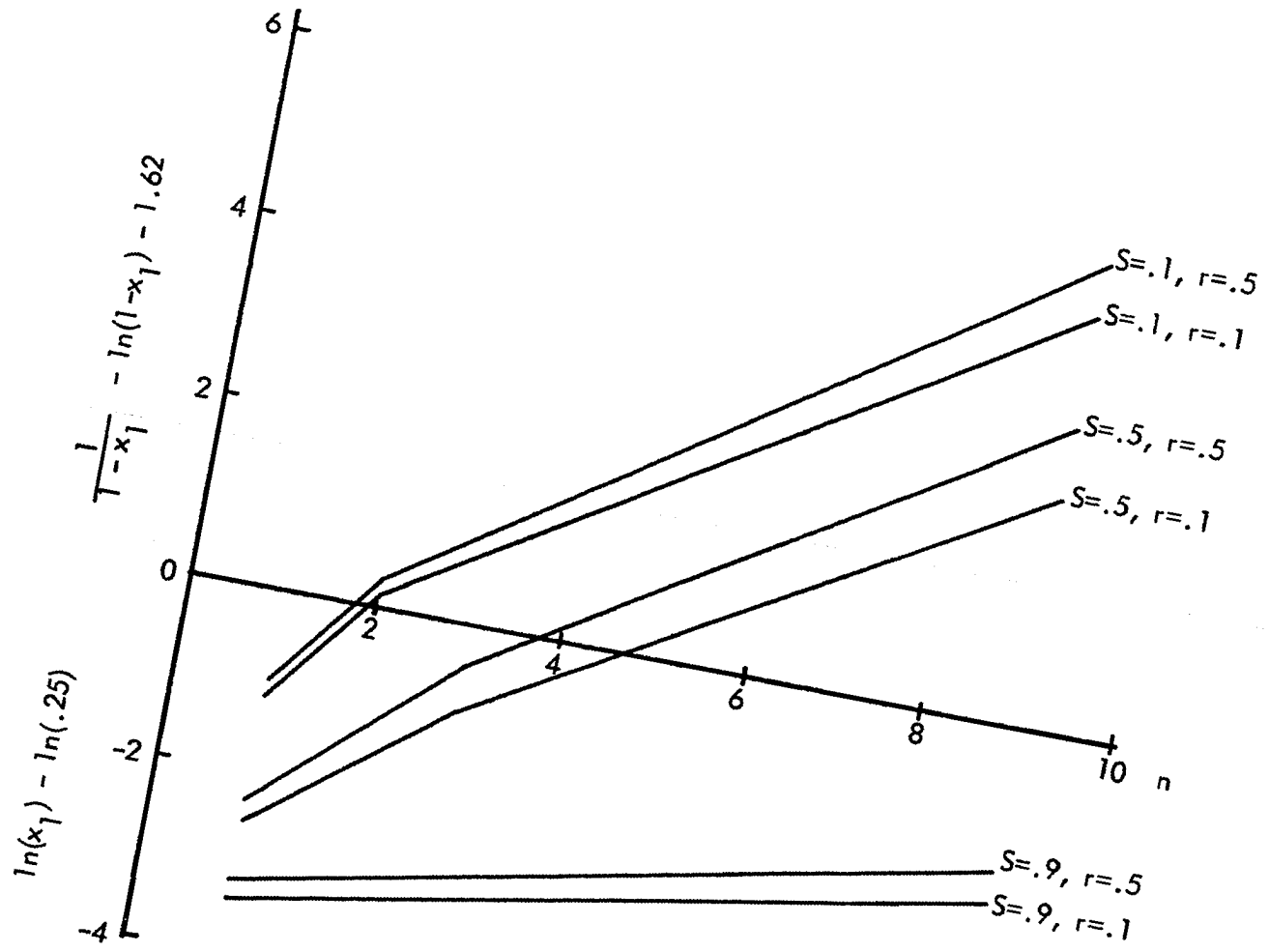


Figure 5. The exact change in the gametic frequency x_1 with generations (n) under different levels of linkage (r) and selection (S). Initial frequency $x_{1,0} = .01$ and $D_0 = 0$. Two Loci-No Variability-No Dominance.

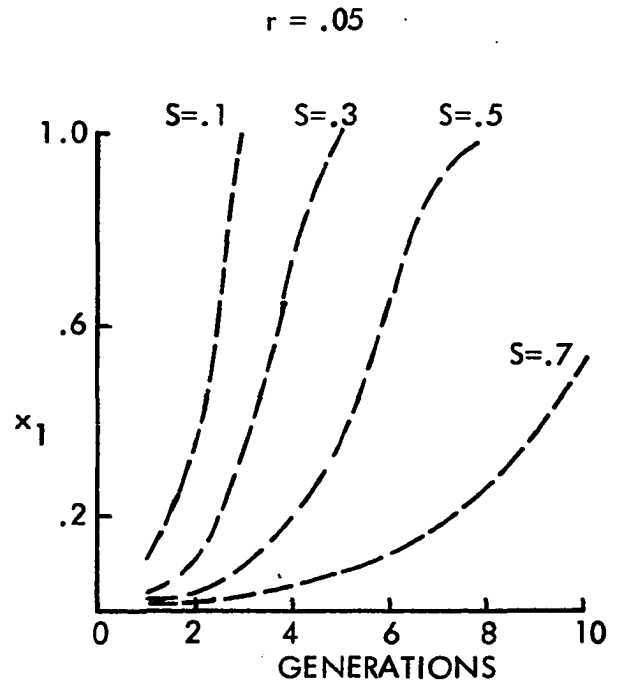
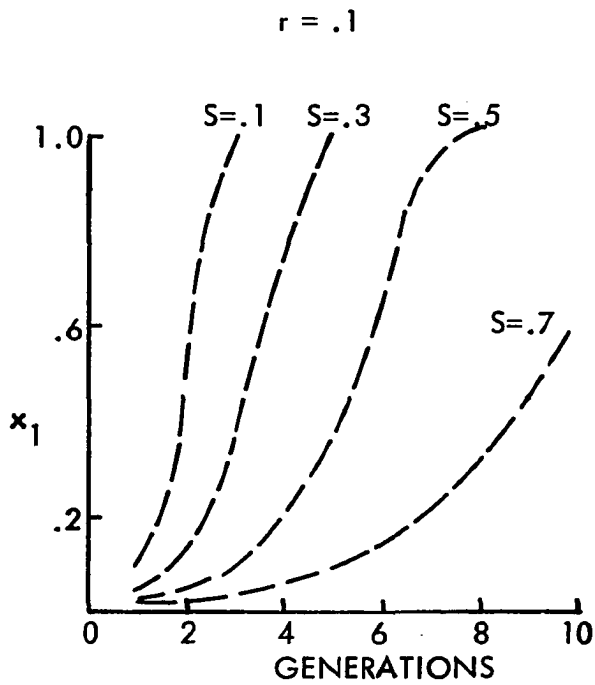
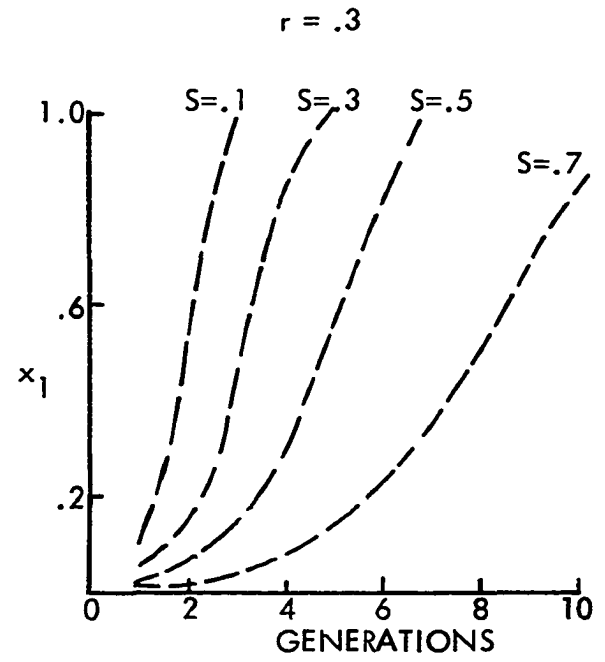
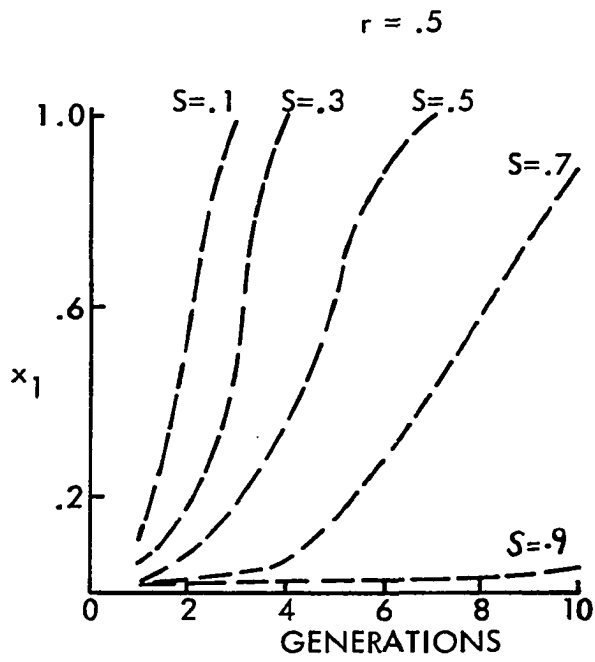


Figure 6. The exact change in gene frequency (p) with generations (n) under different levels of selection (S) and environmental variability (σ^2). Initial frequency $p_0 = .05$. One Locus-Variability.

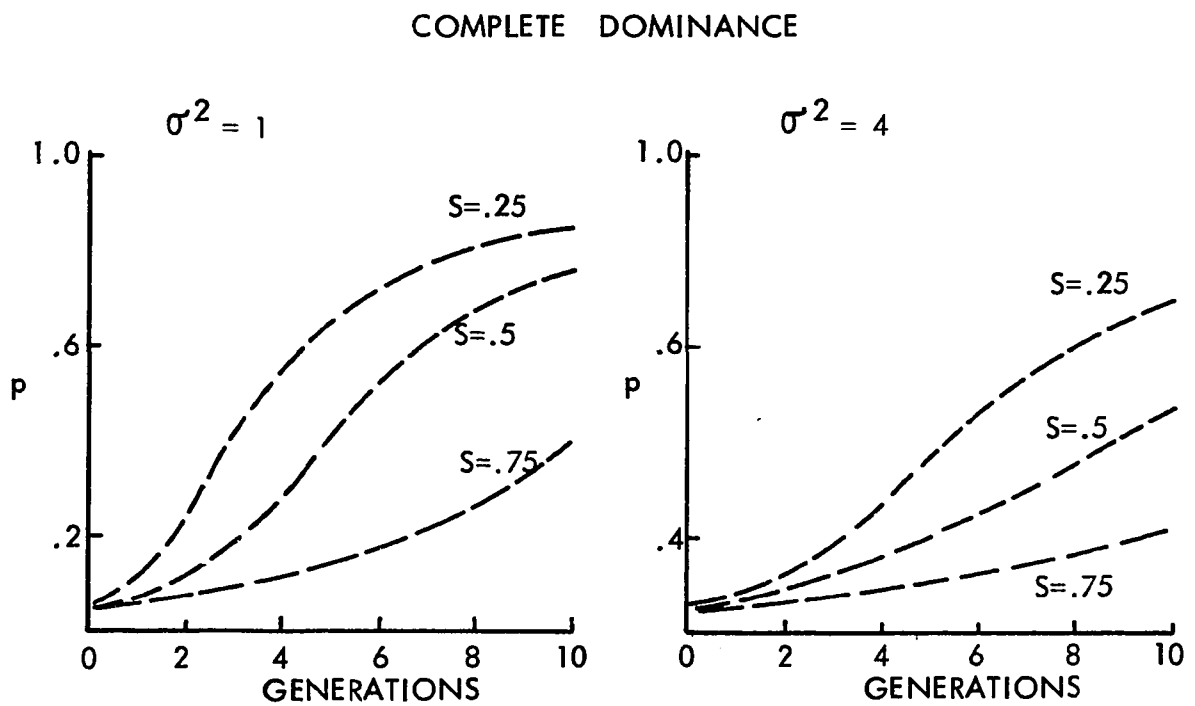
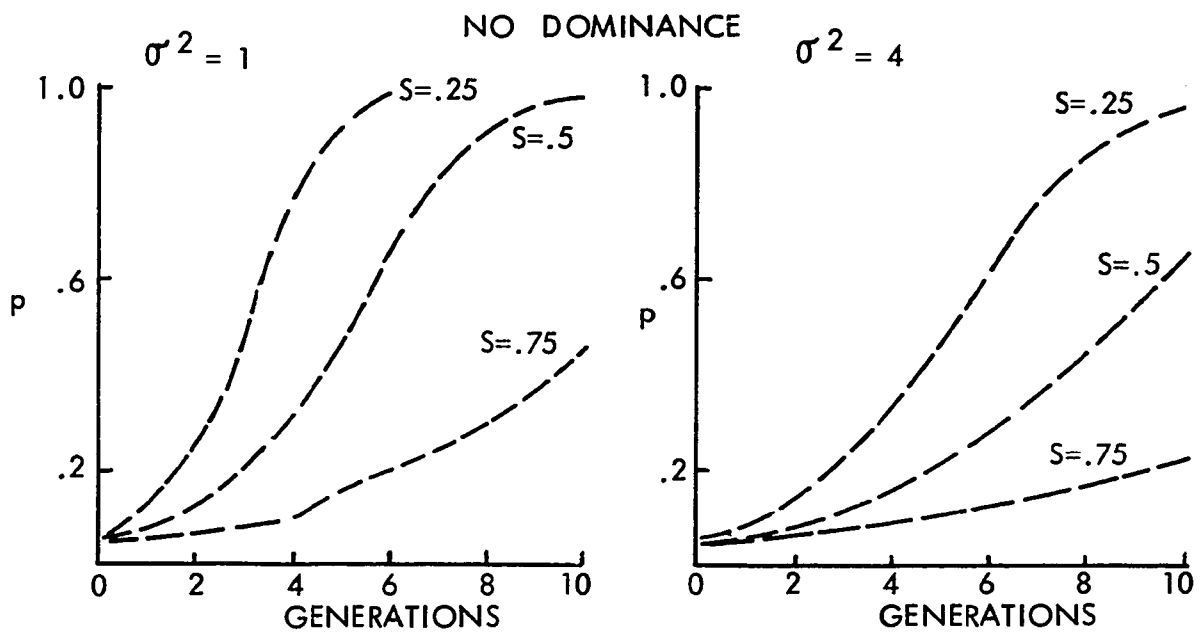
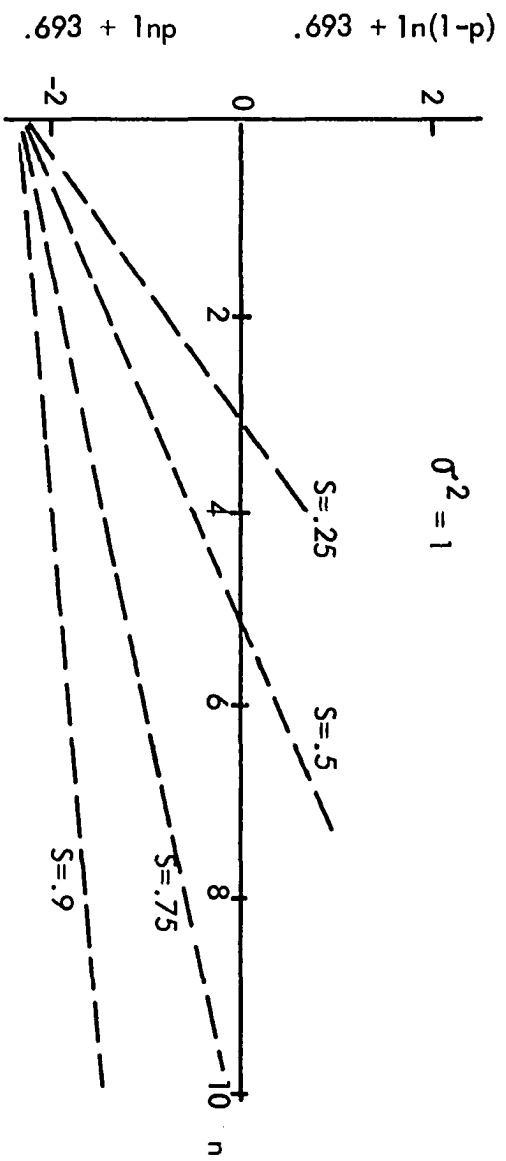
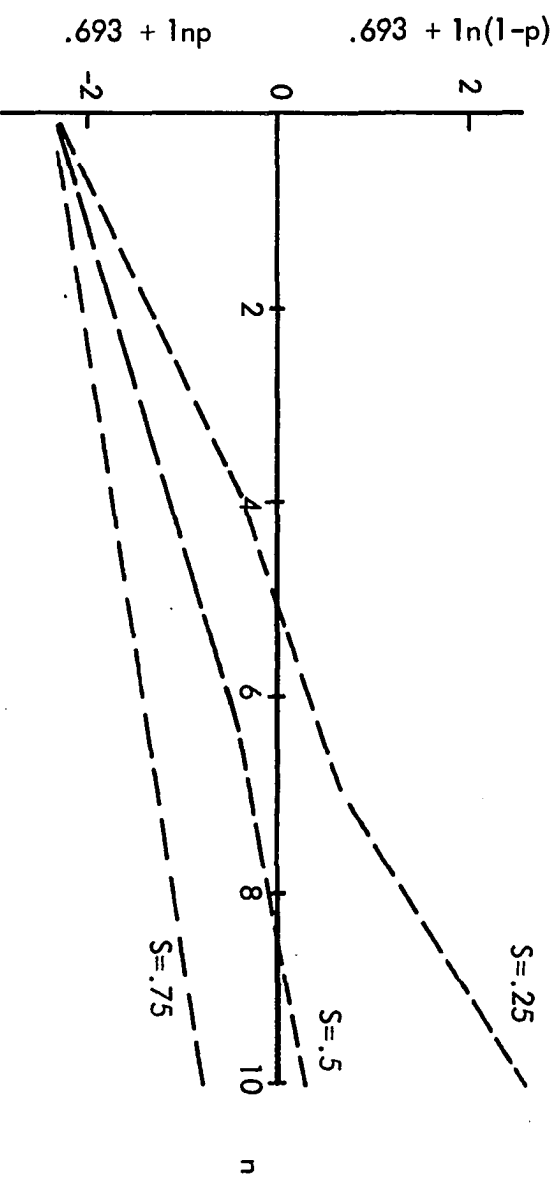


Figure 7. The exact change in the function of the gene frequency (p) with generations (n) under different levels of selection (S) and environmental variability (σ^2). Initial frequency $p_0 = .05$. One Locus-Variability-No Dominance.


 $\sigma^2 = 4$


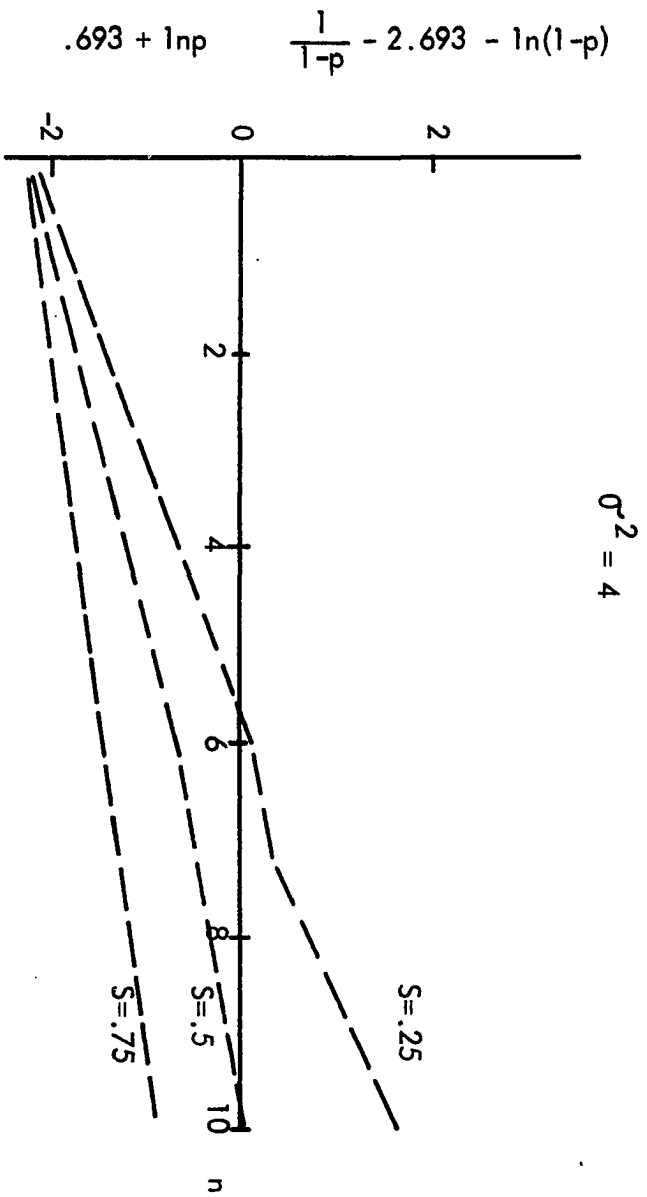
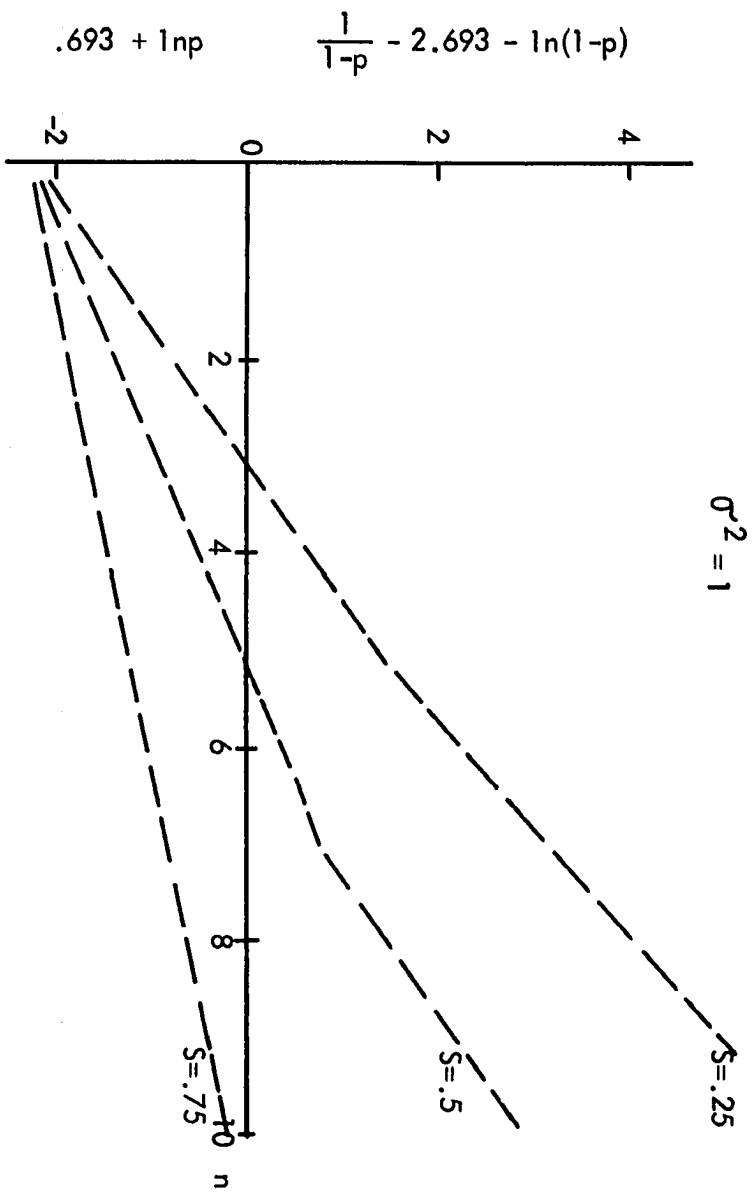


Figure 7 (continued) - Complete Dominance.

Figure 8. The exact change in the gametic frequency x_1 with generations (n) under different levels of selection (S), linkage (r) and environmental variability (σ^2). Initial frequency $x_{1,0} = .01$ and $D_0 = 0$. Two Loci-Variability-No Dominance.

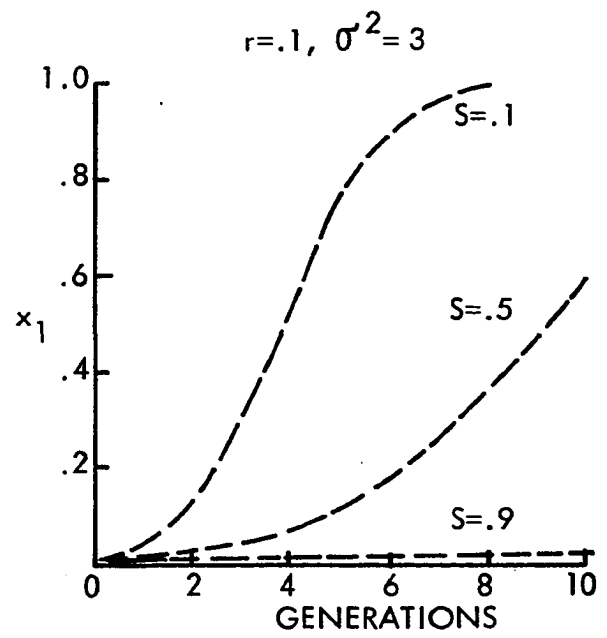
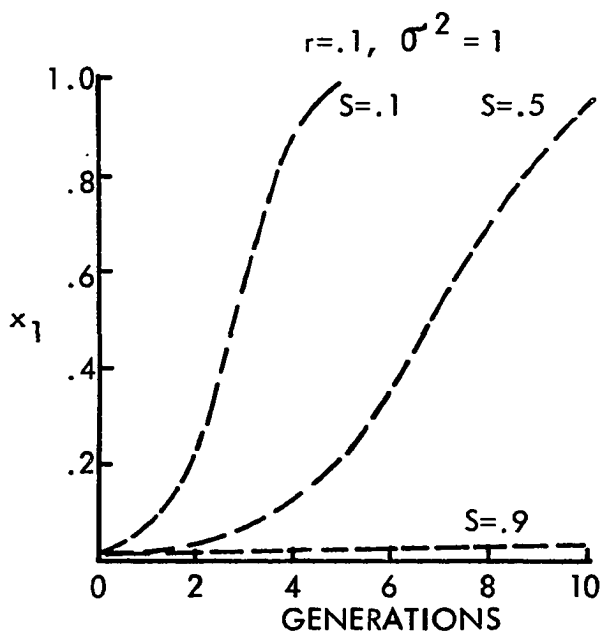
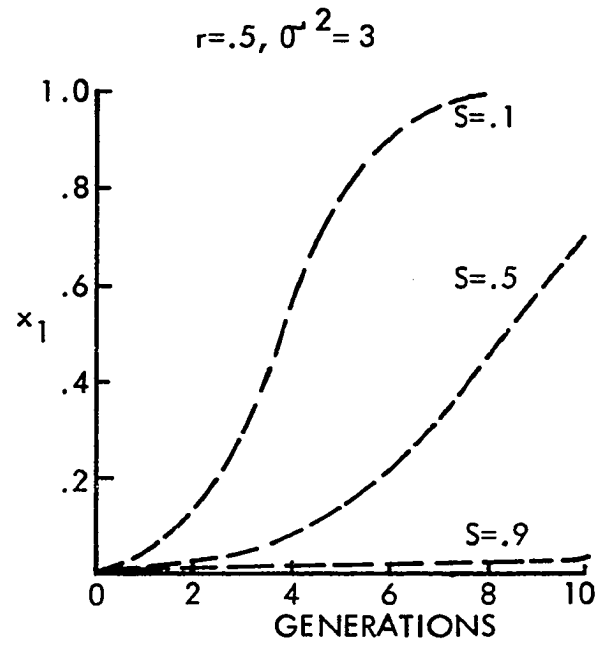
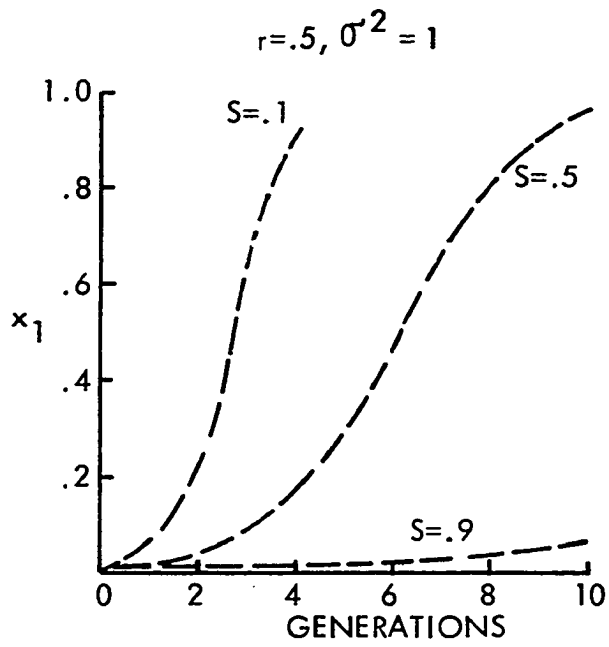


Figure 8 (continued) - Complete Dominance.

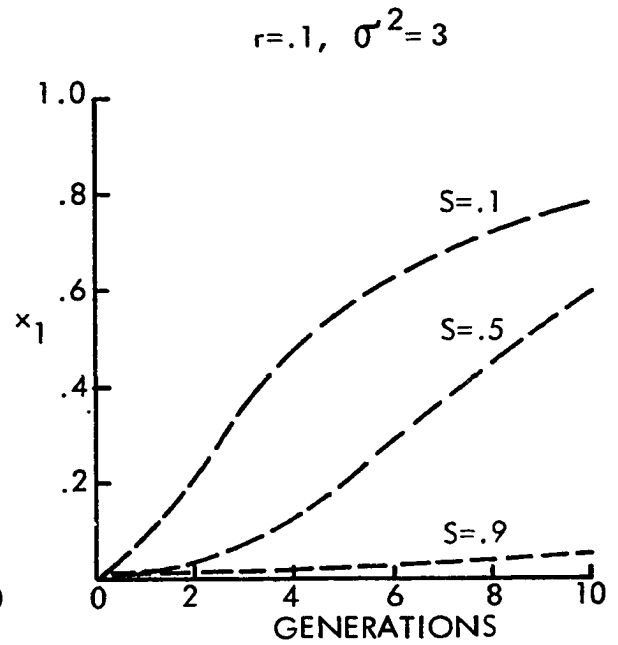
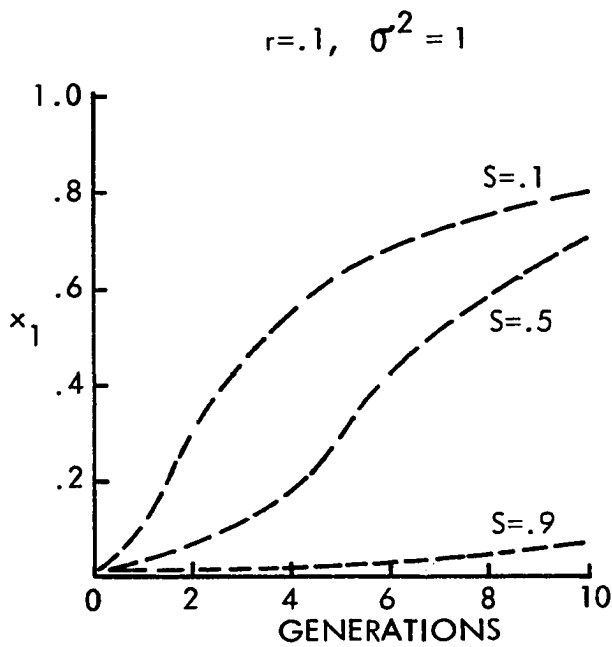
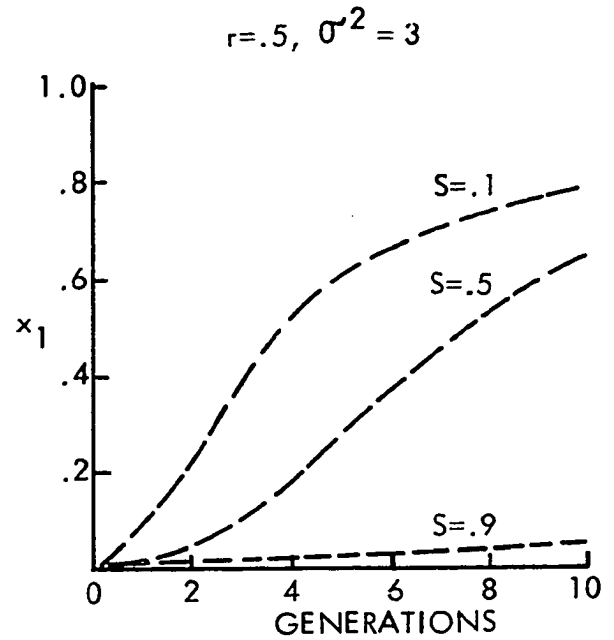
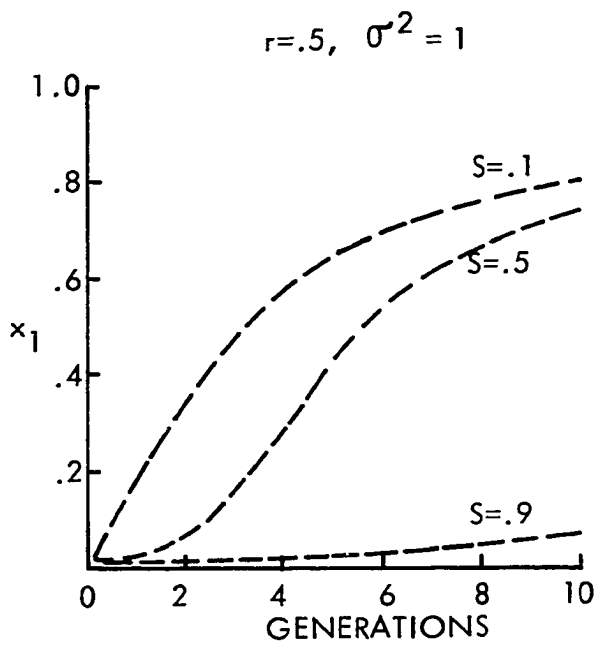
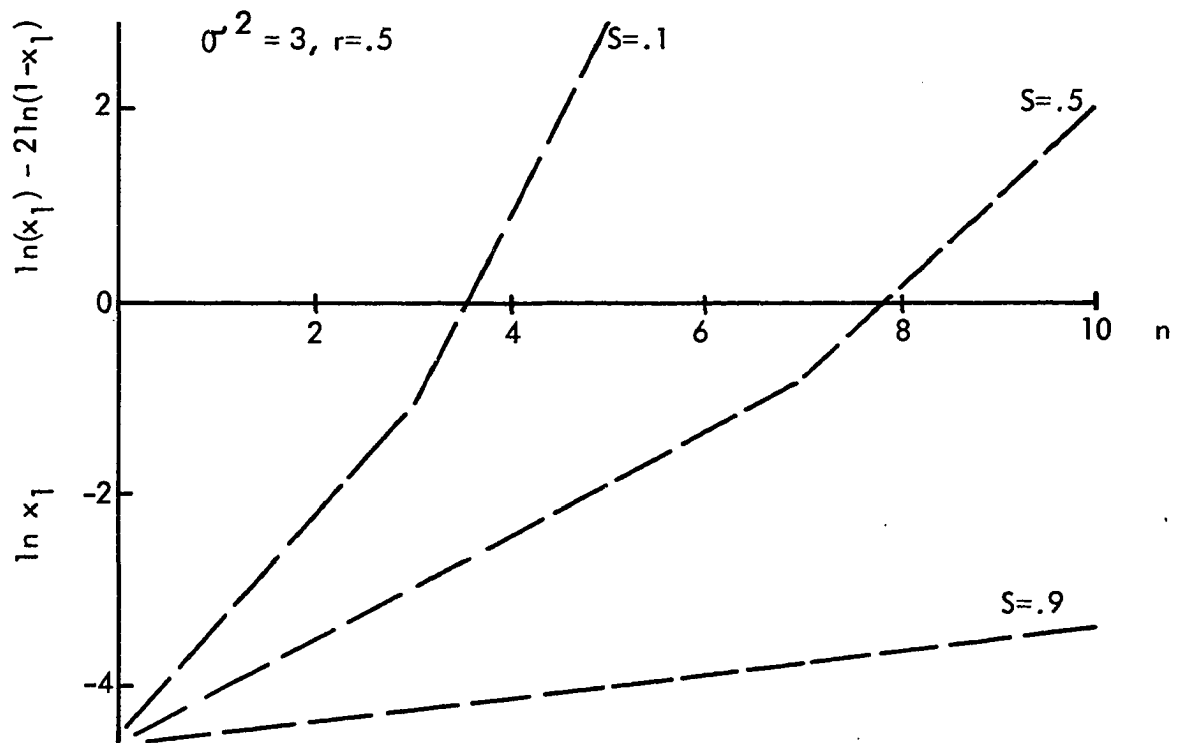
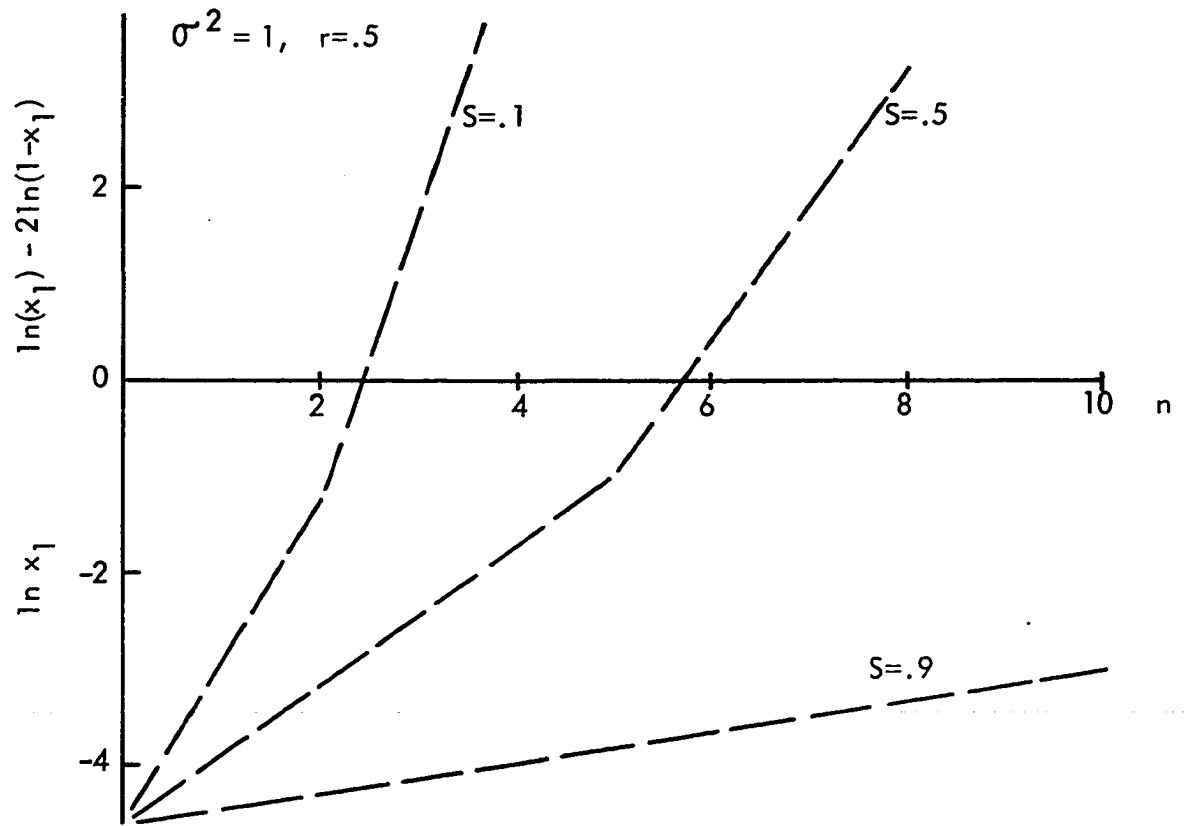


Figure 9. The exact change in the function of the gametic frequency x_1 with generations (n) under different levels of selection (S), linkage (r) and environmental variability (σ^2). Initial frequency $x_{1,0} = .01$ and $D_0 = 0$. Two Loci-Variability-No Dominance.



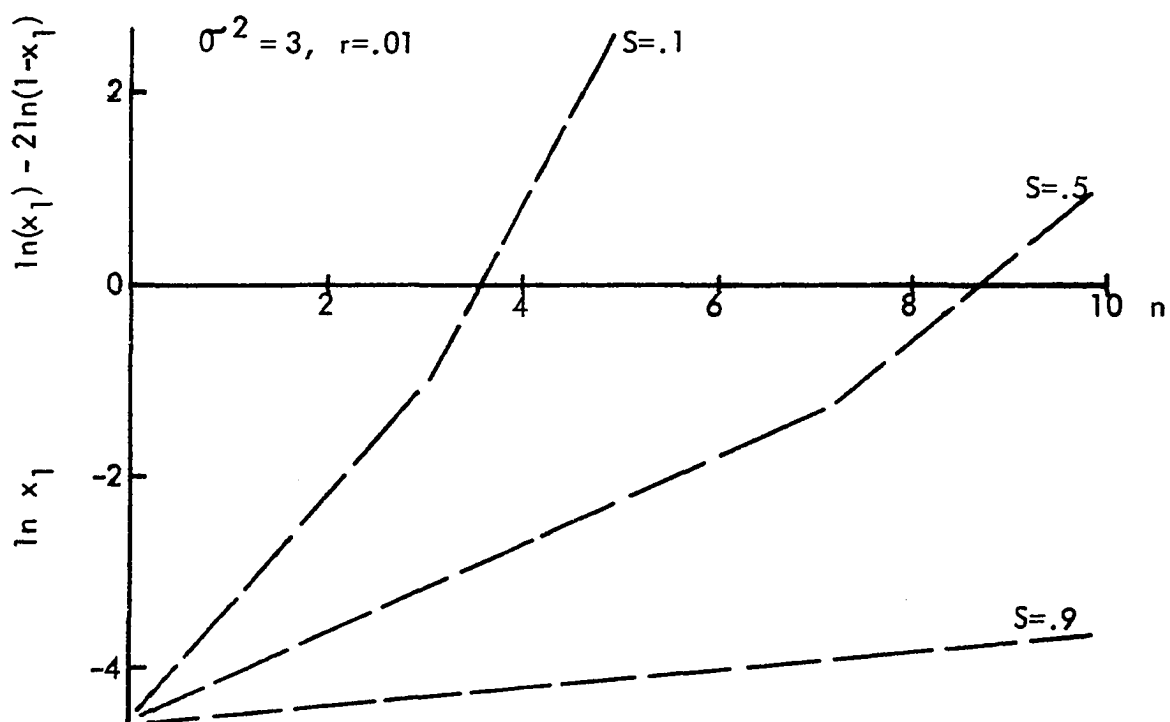
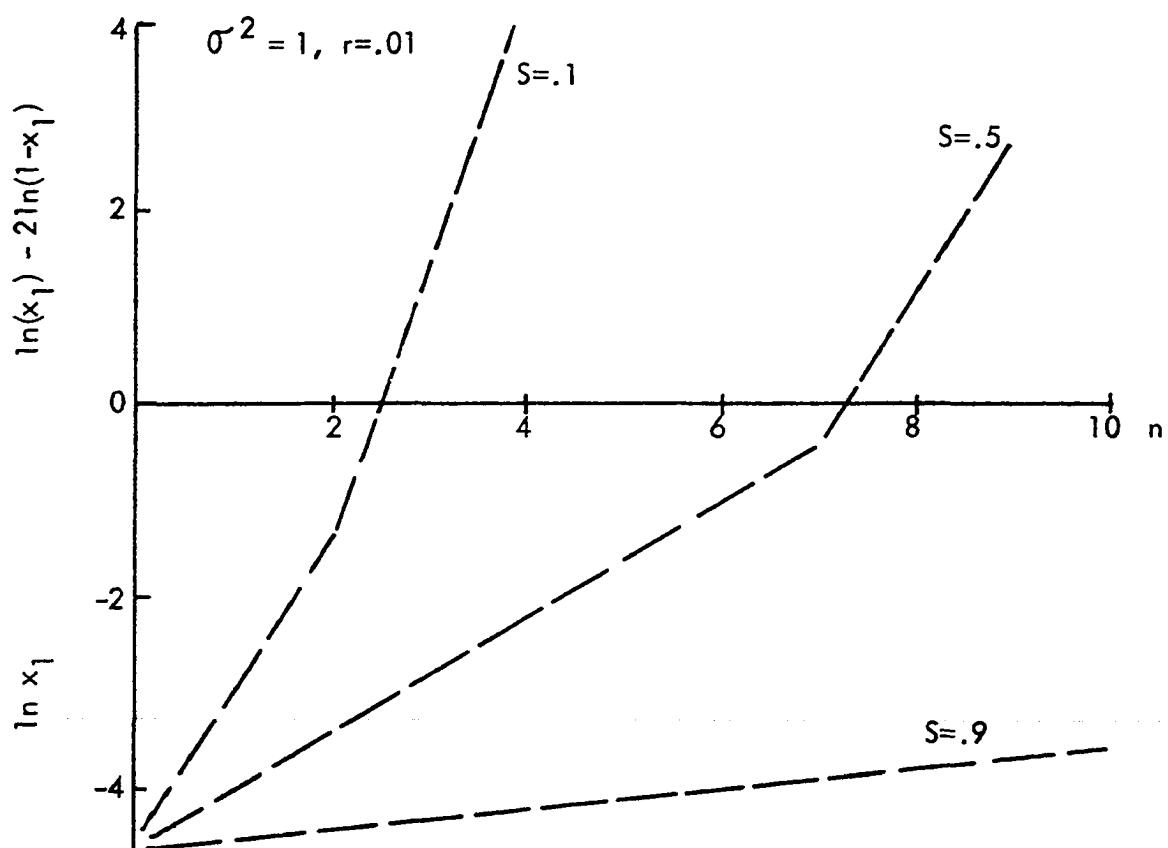
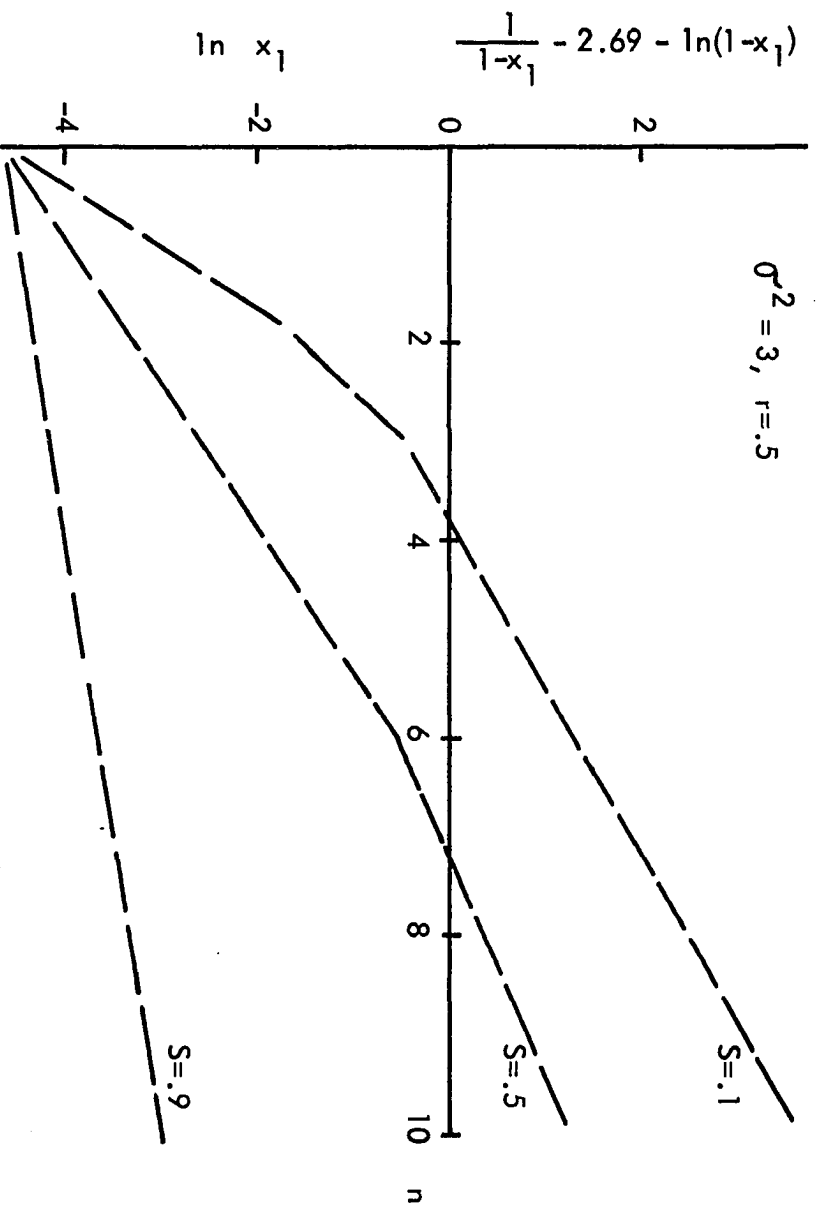
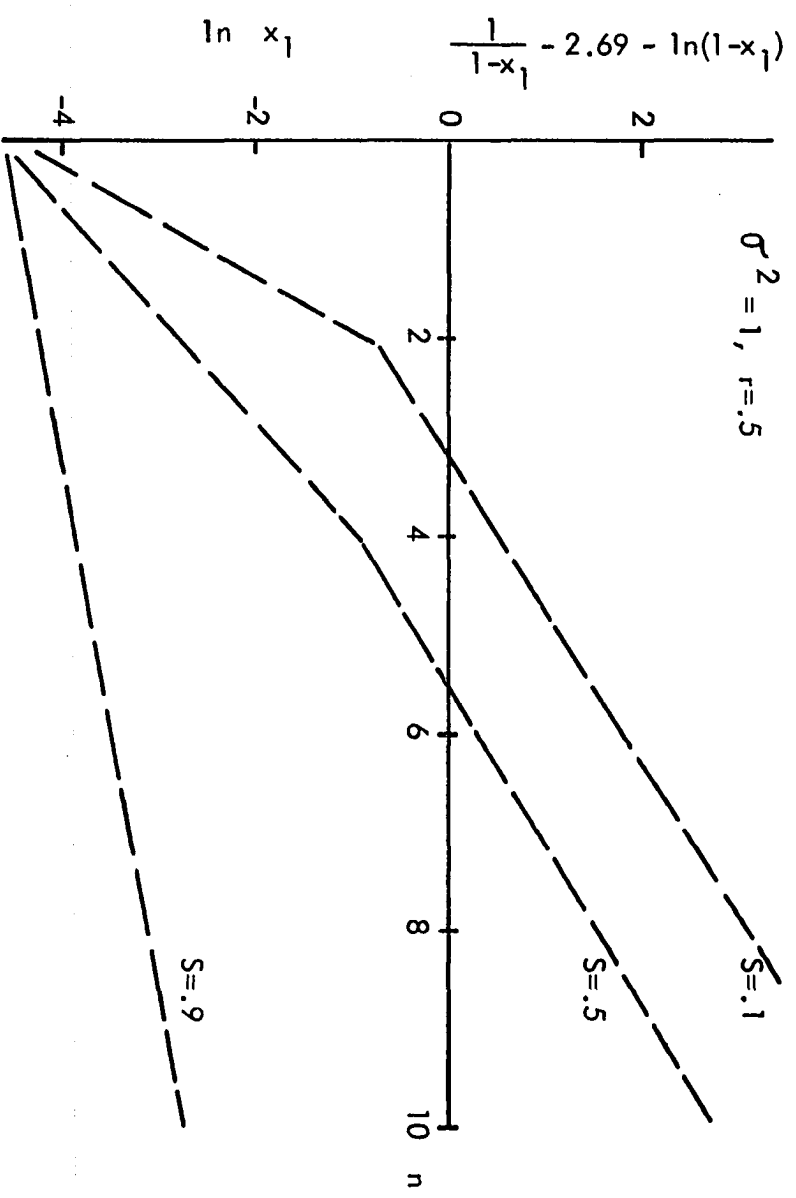


Figure 9 (continued)

Figure 9 (continued) - Complete Dominance.



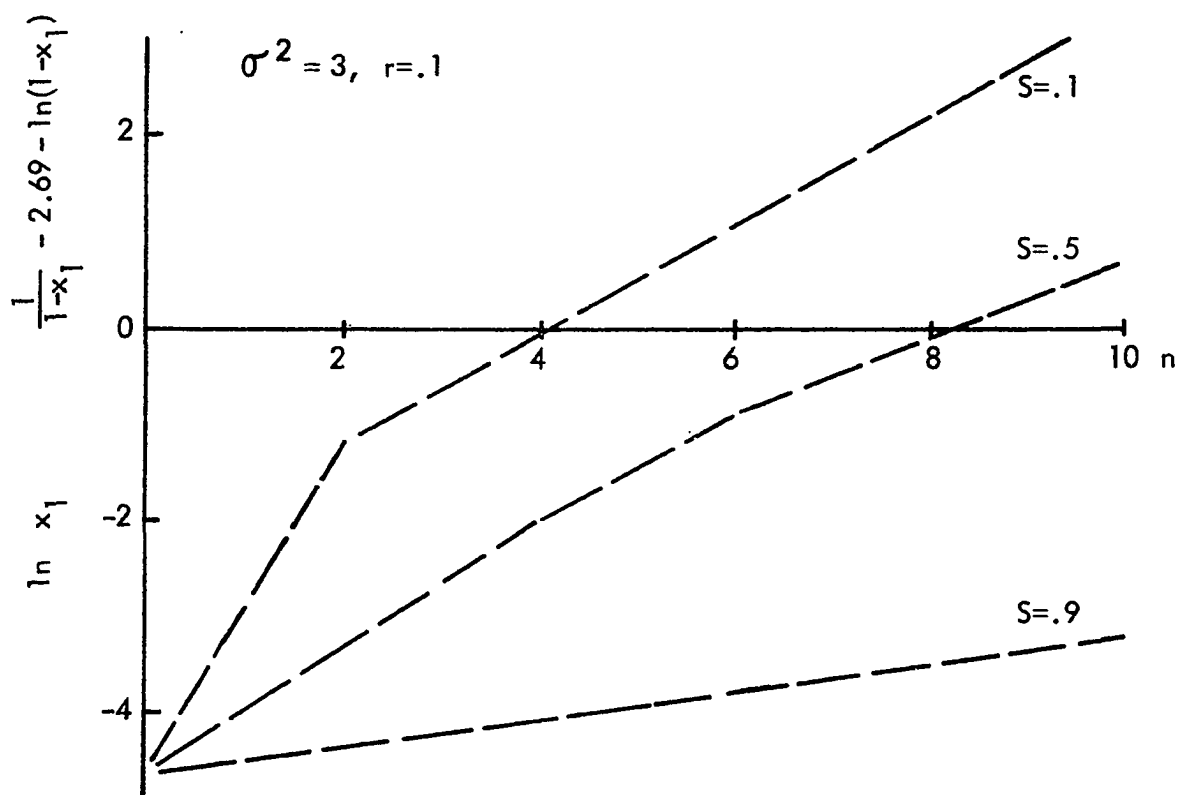
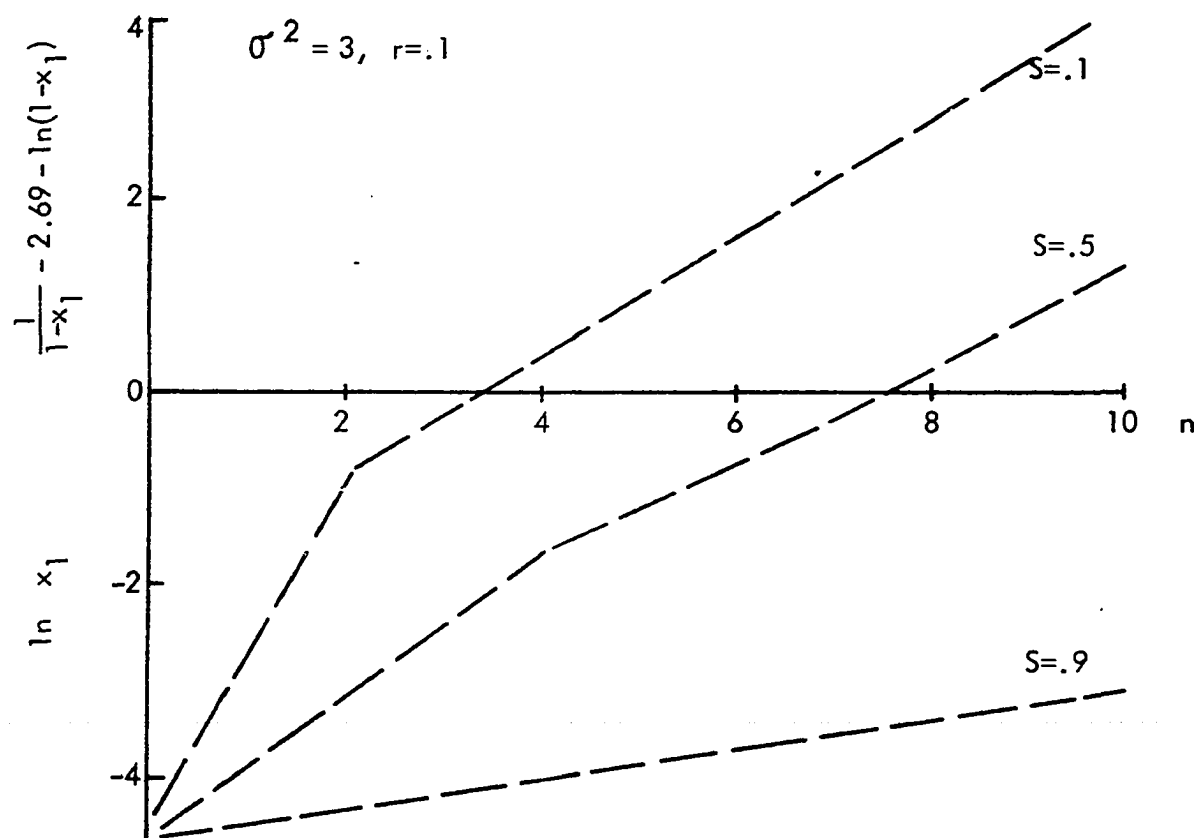


Figure 9 (continued)

APPENDIX B: TABLES

Table 27. Mean phenotypic values for the genetic models.

Model		AA	Aa	aa
1.	No dominance			
	BB	4	3	2
	Bb	3	2	1
	bb	2	1	0
2.	Complete dominance			
	BB	2	2	1
	Bb	2	2	1
	bb	1	1	0
3.	Overdominance			
	BB	0	1	0
	Bb	1	2	1
	bb	0	1	0
4.	Overdominance with epistasis			
	BB	0	0	0
	Bb	0	2	0
	bb	0	0	0

Table 28. Values for the exact slopes β^* of the function of gene frequency and their estimates $\hat{\beta}$ for different models and different levels of selection (S). One Locus - No Variability.

Model	Selection	β	$\hat{\beta}$
No dominance			
First stage	S = .25	1.350	1.386
$(\hat{\beta} = -\ln S)$	= .50	.670	.693
	= .75	.265	.287
	= .90	.100	.106
Complete dominance			
First stage	S = .25	1.33	1.386
$(\hat{\beta} = -\ln S)$	= .50	.67	.693
	= .75	.30	.287
	= .90	.10	.106
Second stage	S = .25	1.20	1.0
$(\hat{\beta} = 1)$	= .50	1.20	1.0
	= .75	1.20	1.0
Overdominance			
First stage	S = .25	1.40	1.386
$(\hat{\beta} = -\ln S)$	= .50	.66	.693
	= .75	.30	.287
	= .90	.10	.106
Second stage	S = .75	.40	.333
$(\hat{\beta} = (1-S)/S)$			

*The values for the exact slope β are obtained from the slopes of the lines in Figure 2.

Table 29. Comparison between the exact slope β^* and their estimates $\hat{\beta}$ for different values of linkage (r) and selection (S). Two Loci - No Variability - No Dominance.

Stage	Selection	Linkage	β	$\hat{\beta}$
Stages 1, 2 and 3 $(\hat{\beta} = (1-S)/S)$	S = .9	r = .50	.16	.11
		= .30	.16	.11
		= .10	.16	.11
		= .05	.16	.11
	S = .7	r = .50	.60	.43
		= .30	.54	.43
		= .10	.45	.43
		= .05	.40	.43
	S = .5	r = .50	1.10	1.00
		= .30	1.00	1.00
		= .10	.86	1.00
		= .05	.78	1.00
	S = .3	r = .50	1.92	2.33
		= .30	1.60	2.33
		= .10	1.30	2.33
		= .05	1.23	2.33
Stage 4 $(\hat{\beta} = \sqrt{(1-S)/S})$	S = .7	r = .50	.33	.33
		= .30	.31	.33
		= .10	.30	.33
		= .05	.30	.33
	S = .5	r = .50	.52	.50
		= .30	.45	.50
		= .10	.45	.50
		= .05	.48	.50
	S = .3	r = .50	.94	.76
		= .30	.78	.76
		= .10	.66	.76
		= .05	.66	.76

*The values for the exact slopes β are obtained from the slopes of the lines in Figure 4.

Table 29 (continued) - Complete Dominance.

Stage	Selection	Linkage	β	$\hat{\beta}$
Stage 1 $(\hat{\beta} = (1-S)/S)$	S = .9	r = .5	.15	.11
		= .1	.16	.11
	S = .7	r = .5	.64	.43
		= .1	.45	.43
	S = .5	r = .5	1.00	1.00
		= .1	.85	1.00
Stage 2 $(\hat{\beta} = 1)$	S = .5	r = .5	.70	1.00
		= .1	.65	1.00
	S = .1	r = .5	.76	1.00
		= .1	.70	1.00

Table 30. Values for the exact slopes β of the function defined by equation (6.32) and their estimates $\hat{\beta}$ given by equation (6.33) for different levels of selection (S), linkage (r) and variability (σ^2).
Two Loci - No Dominance.

Selection	Environmental Variability	Linkage	β	$\hat{\beta}$
S = .1	$\sigma^2 = 1$	r = .01	2.30	2.0
		= .10	2.35	2.0
		= .50	2.45	2.0
	$\sigma^2 = 3$	r = .01	1.38	1.2
		= .10	1.35	1.2
		= .50	1.40	1.2
S = .5	$\sigma^2 = 1$	r = .01	.90	.40
		= .10	.85	.40
		= .50	.85	.40
	$\sigma^2 = 3$	r = .01	.30	.23
		= .10	.32	.23
		= .50	.33	.23
S = .9	$\sigma^2 = 1$	r = .01	.11	.22
		= .10	.18	.22
		= .50	.20	.22
	$\sigma^2 = 3$	r = .01	.10	.13
		= .10	.11	.13
		= .50	.12	.13

Table 31. Values for the exact slope β of the linear relation of $\ln x_1$ with generations and their estimates $\hat{\beta}$ for different levels of selection (S), linkage (r) and variability (σ^2). Two Loci - No Dominance.

Selection	Environmental Variability	Linkage	β	$\hat{\beta}$
S = .5	$\sigma^2 = 1$	r = .01	.50	1.0
		= .10	.65	1.0
		= .50	.72	1.0
	$\sigma^2 = 3$	r = .01	.50	1.0
		= .10	.56	1.0
		= .50	.57	1.0
S = .9	$\sigma^2 = 1$	r = .01	.13	.11
		= .10	.16	.11
		= .50	.18	.11
	$\sigma^2 = 3$	r = .01	.10	.11
		= .10	.13	.11
		= .50	.14	.11

$$*\hat{\beta} = (1-S)/S.$$

Table 32. Values for the exact slopes β of the functions of the gametic frequency x_1 and their estimates $\hat{\beta}$ for different levels of selection (S), linkage (r) and environmental variability (σ^2). Two Loci - Variability. Complete Dominance.

Selection	Environmental Variability	Linkage	β	$\hat{\beta}$
Stage 1				
S = .5	$\sigma^2 = 1$	r = .10	.76	1.0
		= .50	.95	1.0
	$\sigma^2 = 3$	r = .10	.65	1.0
		= .50	.73	1.0
S = .9	$\sigma^2 = 1$	r = .10	.13	.11
		= .50	.20	.11
	$\sigma^2 = 3$	r = .10	.12	.11
		= .50	.14	.11
Stage 2				
S = .1	$\sigma^2 = 1$	r = .10	.66	1.0
		= .50	.70	1.0
	$\sigma^2 = 3$	r = .10	.55	1.0
		= .50	.60	1.0
S = .5	$\sigma^2 = 1$	r = .10	.60	1.0
		= .50	.65	1.0
	$\sigma^2 = 3$	r = .10	.50	1.0
		= .50	.50	1.0